PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of

Valerie CHEYNET-SAUVION et al.

Group Art Unit: 1655

Application No.: 09/402,131

Examiner: B. Sisson

Filed: December 8, 1999

Docket No.: 104458

For: RNA-DEPENDENT RNA POLYMERASE FUNCTIONING PREFERABLY ON RNA MATRIX AND PROMOTER-DEPENDENT TRANSCRIPTION PROCESS WITH SAID RNA-DEPENDENT RNA POLYMERASE

DECLARATION UNDER 37 C.F.R. §1.132

Director of the U.S. Patent and Trademark Office
Washington, D.C. 20231

Sir:

I, William T. McAllister, a citizen of the United States of America, hereby declare and state:

1. My qualifications are set forth in the attached curriculum vitae.

2. I am an inventor of the above-identified patent application.

3. I am familiar with its contents and the contents of the pending Office Action and the accompanying Amendment After Final Rejection under 37 C.F.R. 1.116.

4. I have read and understand the attached references, and believe that the teachings of the attached references represent the state of the art at the time the application was filed. I believe that the specification as originally filed fully enabled one skilled in the art to make and use the invention as recited in the claims as amended by the accompanying Amendment After Final Rejection for the reasons set forth below.

5. One skilled in the art would have understood that RNA polymerases encoded by bacteriophage T7 and its relatives have many common structural and functional features.

The promoter sequences recognized by these phage polymerases, such as T7, T3, K11, SP6 and BA14 phage polymerases, all share a common 23 bp consensus sequence between nucleotides -17 to +6 (see e.g., McAllister, *Cellular and Molecular Biology Research* (1993) 39: 385-391). Because of the structural and functional similarities of the RNA polymerase encoded by these bacteriophages those skilled in the art refer to these RNA polymerases as "T7-like RNA polymerases" and to the phages as "T7-like bacteriophages." (See e.g., Chamberlin et al. (copy submitted with June 5, 2001, Amendment) at page 88, first paragraph, and page 89, Heading 11). Thus, referring to the RNA polymerase as "T7-like phage polymerase" would clearly have been understood by one skilled in the art as referring to the RNA polymerases encoded by T7 and its related phages.

6. In addition to recognizing this consensus sequence within the transcription promoter sequence, T7-like phage RNA polymerases also have a common organization. It is known in the art that these RNA polymerases consist of a single subunit (see e.g., Severinov, PNAS, (2000) 98: 5-7; Tahirov et al., Nature (2002) 420:43-50; and Yin et al., Science (2002) 298: 1387-1395). These references show that there are two "families" of DNA-dependent RNA polymerases that are recognized in the art. One family of polymerases encompasses the T7-like phage polymerases, which consist of a single subunit, while the second family of RNA polymerases covers bacterial and eukaryotic RNA polymerases, which consist of multiple subunits. As described in the specification at, for example, page 4, line 30 to page 53, the T7-like phage polymerases are an art-recognized class of very homologous enzymes. This is also supported by the discussion in Chamberlin et al., from page 89 to 91. Furthermore, Severinov, Tahirov et al. and Yin et al. further demonstrate and confirm the accuracy of this grouping of phage RNA polymerases. Thus, those skilled in the art would have recognized that the T7-like phage polymerases are a closely related group of RNA polymerases.

7. The ability to synthesize a polymer of nucleotides is conferred by the active site of the enzyme, which is highly conserved among the T7-like phage polymerases. An alignment of the RNA polymerases from exemplary T7, T3, SP6 and K11 RNA polymerases is attached. The alignment shows that the amino acid sequence from residue 620 to about 640 is highly conserved across the different types of T7-like phage polymerases.

8. The conserved amino acid sequence in this region of the T7-like phage RNA polymerases would also have suggested to one skilled in the art that similar changes made in the phage polymerases would result in the same or similar mutant phenotypes. Thus, one skilled in the art would have reasonably expected that a mutation at R627 in the manner described in the specification with respect to T7 RNA polymerase would result in a similar mutant phenotype as that of the T7 RNA polymerase in other T7-like phage polymerases. It is a generally accepted and routine practice among those skilled in the art to compare the amino acid sequence of related proteins to localize areas of importance and interest.

9. The skilled artisan would have determined an appropriate mutagenesis strategy based on the comparison of the amino acid sequences and structures. Thus, there would have been no need for the skilled artisan to examine multiple mutations at every possible position within the protein as asserted in the Office Action. The demonstration of one mutant of T7 RNA polymerase activity within a highly conserved region of the amino acid sequence shared by the T7-like phage RNA polymerases would have been expected to yield similar results in other T7-like phage polymerases. Thus, no undue experimentation would have been necessary to practice the claimed invention with various alternative T7-like phage polymerases.

10. The specification describes the modification of a T7 RNA polymerase at residue R627. As a result of this modification, the RNA dependent RNA polymerase activity of T7 RNA polymerase is greatly enhanced. As discussed above, this particular residue lies

within the highly conserved region, between amino acid residues 620 to about 640, that is shared by the T7-like phage RNA polymerases. Thus, one skilled in the art would have expected that the same or similar modification in the highly conserved regions within a different, but related, T7-like phage RNA polymerase would also enhance the RNA dependent RNA polymerase activity in the related T7-like phage RNA polymerase.

11. Thus, in view of the attached references, the specification as filed provides a fully enabling disclosure for the claimed invention. One skilled in the art would not have required further guidance or examples, nor would undue experimentation have been required to practice the claimed invention beyond what is disclosed in the specification.

12. I hereby declare that all statements made herein of my own knowledge are true, and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine and/or imprisonment under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing therefrom.

Date: _____ _____

William T. McAllister

CURRICULUM VITAE

WILLIAM T. MCALLISTER

PRESENT POSITION:

Professor and Chairman
Department of Microbiology and Immunology
SUNY Health Sciences Center at Brooklyn, Brooklyn, NY

PERSONAL DATA:

Date of Birth: April 25, 1944
Three children: Elliot, Suzanne, Robert

EDUCATION:

1961-66	Lehigh University, Bethlehem, PA B.A. (Biology)
1966-67	Department of Microbial and Molecular Biology University of Pittsburgh, Laboratory of Dr. D. MacDonald Green
1967-70	Department of Biochemistry University of New Hampshire, Ph.D. (Biochemistry) Laboratory of Dr. MacDonald Green (genetics and biology of *B.subtilis* phages)
1970-72	Institute for Molecular Genetics Heidelberg University, Germany NIH Postdoctoral Fellow Laboratory of Dr. E.K.F. Bautz (molecular genetics, control of transcription)

PROFESSIONAL EXPERIENCE:

1966-67	Teaching Assistant, Department of Microbial and MolecularBiology University of Pittsburgh
1967-70	Research Assistant/Predoctoral Trainee, Department of Biochemistry, University of New Hampshire
1970-72	NIH Postdoctoral Fellow, Institute for Molecular Genetics Heidelberg University
1972-73	Research Associate, Institute for Molecular Genetics Heidelberg University

1973-79	Assistant Professor, Department of Microbiology, College of Medicine and Dentistry of New Jersey-Rutgers Medical School, Piscataway, New Jersey
1979-85	Associate Professor, Department of Microbiology, University of Medicine and Dentistry of New Jersey-Rutgers Medical School, Piscataway, New Jersey
1985-86	Professor, Department of Microbiology, University of Medicine and Dentistry of New Jersey-Rutgers Medical School, Piscataway, New Jersey
1986-	Professor and Chairman, Department of Microbiology and Immunology, State University of New York Health Sciences Center at Brooklyn, Brooklyn, NY

HONORS AND AWARDS:

Chancellor's Award for Excellence in Scholarship and Creative Activities, State University of New York, 2002

National Lecturer, American Society for Microbiology, 1986-87

Excellence in Teaching Award, Foundation of UMDNJ, Rutgers Medical School, 1980

Councilor, Harvey Society, 1987-1992

Chair, Division M (Bacteriophages), American Society for Microbiology, 1991-1992

Councilor, Association of Medical School Microbiology Chairs, 1994-1997

Faculty GEM Award for Outstanding Research, Alumni Association, College of Medicine, Downstate Medical Center, Brooklyn, NY 2001

RESEARCH SUPPORT:

Current:

NIH GM38147 "RNA polymerase structure and function", 12/01/98-11/30/02, Principal Investigator, Total project, $1,748,299; current year, $421,862.

Howard Hughes Medical Institute "Structure-function relationships of bacteriophage T7 RNA polymerase", 7/1/95-6/30/01. This is a collaborative project with Dr. Sergei Kochetkov, Moscow; I am the Lead Collaborating Scientist. Direct costs (to US laboratory): total project, $14,000; current year, $2,800.

Prior:

bioMerieux, S.A. "RNA polymerases with altered specificities", 7/1/93-6/30/98, Principal Investigator.

Life Technologies, Inc. "Development of phage RNA polymerase-based expression system", 7/1/92-6/30/94, Principal Investigator.

NIH-GM21783 "Regulation of viral gene expression", 07/01/74-6/30/88, Principal Investigator.

NIH AM-28561, "Folate binders, hematopoiesis, and cell replication", 6/1/87-5/31/92, Co-investigator.

New Jersey Commission on Cancer Research, 86-187-CCR, "Studies of recombination in papillomavirus-transformed cells", 07/01/85-6/30/87, Principal Investigator

NSF 412-6200A, "Support for an international workshop on gene organization and expression in bacteriophages", 7/1/88-6/30/89, Principal Investigator

NSF MCB9602092, "Support of an International Workshop on Macromolecular Interactions in Bacteriophages", 7/1/96-6/30/97, Principal Investigator

Pharmacia P-L Biochemicals, Inc, "Cloning and expression of the bacteriophage SP6 RNA polymerase gene", 01/01/87-12/31/90, Principal Investigator

Life Technologies, Inc., "Development of phage RNA polymerase-based expression systems", 01/01/91-12/31//94, Principal Investigator

Biotechnology Research and Development Corporation "Development of a plant expression system based upon phage T3 RNA polymerase", 4/1/92-8/31/93, Principal Investigator

PUBLICATIONS

Articles:

1. McAllister, W.T. (1970). Bacteriophage infection: Which end of the SP82G genome goes in first? J. Virology 5:194-198.

2. McAllister, W.T., and Green, D.M. (1972). Bacteriophage SP82G inhibition of an intracellular deoxyribonucleic acid inactivation process in Bacillus subtilis. J. Virology 10:51-59.

3. McAllister, W.T., and Green, D.M. (1973). Effects of the decay of incorporated radioactive phosphorus on the transfer of the bacteriophage SP82G genome. J. Virology 12: 300-309.

4. Dunn, J.J., McAllister, W.T., and Bautz, E.K.F. (1972). In vitro transcription of T3 DNA by E. coli and T3 RNA polymerases. Virology 48:112-125.

5. Dunn, J.J., McAllister, W.T., and Bautz, E.K.F. (1972). Transcription in vitro of T3 DNA by E. coli and T3 RNA polymerases II. Analysis of the products in a cell-free protein-synthesizing system. Eur. J. Biochem. 29:500-508.

6. Bautz, E.K.F., McAllister, W.T., and Kupper, H. (1972). RNA polymerase of bacteriophage T3. Studia Biophysics 31/32:7-14.

7. McAllister, W.T., Kupper, H., and Bautz, E.K.F. (1973). Kinetics of transcription by the bacteriophage T3 RNA polymerase in vitro. Eur. J. Biochem. 34:489-501.

8. Kupper, H., McAllister, W.T., and Bautz, E.K.F. (1973). Comparison of E. coli and T3 RNA polymerases: Differential inhibition of transcription by various drugs. Eur. J. Biochem. 38:581-586.

9. Bautz, E.K.F., McAllister, W.T., Kupper, H., Beck, E., and Bautz, F.A. (1974). Initiation of transcription by RNA polymerases of E. coli and phage T3. Adv. Exptl. Med. and Biol. 44:7-21.

10. McAllister, W.T., and Barrett, C.L. (1977). Hybridization mapping of restriction fragments from the early region of bacteriophage T7 DNA. Virology 82:275-287.

11. McAllister, W.T., and McCarron, R.J. (1977). Hybridization of the in vitro products of bacteriophage T7 RNA polymerase to restriction fragments of T7 DNA. Virology 82:288-298.

12. McAllister, W.T., and Barrett, C.L. (1977).Roles of the early genes of bacteriophage T7 in shutoff of host macromolecular synthesis. J. Virology 23:543-553.

13. McAllister, W.T., and Barrett, C.L. (1977). Superinfection exclusion by bacteriophage T7. J. Virology 24:709-711.

14. McCarron, R.J., Cabrera, C.V., Esteban, M., McAllister, W.T., and Holowczak, J.A. (1978). Structure of vaccinia DNA: Analysis of the viral genome by restriction endonucleases. Virology 86:88-101.

15. Cabrera, C.V., Esteban, M., McCarron, R., McAllister, W.T., and Holowczak, J.A. (1978). Vaccinia virus transcription: Hybridization of mRNA to restriction fragments of vaccinia DNA. Virology 86:102-114.

16. McCarron, R.J., and McAllister, W.T. (1978). Effect of ribosomal loading on the structural stability of bacteriophage T7 early messenger RNAs. Biochem. Biophys. Res. Commun. 80:789-796.

17. McAllister, W.T., and Wu, H-L. (1978). Regulation of transcription of the late genes of bacteriophage T7. Proc. Natl. Acad. Sci. (USA) 75:804-808.

18. Bailey, J.N., Dembinski, D.R., and McAllister, W.T. (1980). Derivation of a restriction map of bacteriophage T3 DNA and comparison with the map of bacteriophage T7 DNA. J. Virol. 35:176-183.

19. McAllister, W.T., and Carter, A.D. (1980). Regulation of promoter selection by the bacteriophage T7 RNA polymerase in vitro. Nuc. Acids Res. 8:4821-4837.

20. Bailey, J.N. and McAllister, W.T. (1980). Mapping of promoter sites utilized by T3 RNA polymerase on T3 DNA. Nuc. Acids Res. 8:5071-5088.

21. Carter, A.D., Morris, C.E., and McAllister, W.T. (1981). A revised transcription map of the late region of bacteriophage T7 DNA. J. Virol. 37:636-642.

22. McCarron, R.J., and McAllister, W.T. (1981). Effect of alterations in reaction conditions on vaccinia virus transcription in vitro. Virology, 113:392-396.

23. McAllister, W.T., Morris, C.E., Studier, F.W., and Rosenberg, A. (1981). Utilization of T7 late promoters in recombinant plasmids by the T7 RNA polymerase in vivo. J. Mol. Biol. 153:527-544.

24. Carter, A.D., and McAllister, W.T. (1981). Sequences of three class II promoters for the bacteriophage T7 RNA polymerase. J. Mol. Biol. 153:825-830.

25. Jolliffe, L.K., Carter, A.D., and McAllister, W.T. (1982). Identification of a potential control region in bacteriophage T7 late promoters. Nature 299:653-656.

26. Bailey, J.N., Klement, J.F., and McAllister, W.T. (1983). Relationship between promoter structure and template specificities exhibited by the bacteriophage T3 and T7 RNA polymerases. Proc. Nat. Acad. Sci. (USA) 80:2814-2818.

27. McGraw, N.J., Bailey, J.N., Cleaves, G.R., Dembinski, D.R., Gocke, C.R., Jolliffe, L.K., MacWright, R.W. and McAllister, W.T. (1985). Sequence and analysis of the gene for bacteriophage T3 RNA polymerase. Nuc. Acids Res. 13:6753-6766.

28. Brown, J.E., Klement, J.F. and McAllister, W.T. (1986). Sequences of three promoters for the bacteriophage SP6 RNA polymerase. Nucl. Acids Res. 14:3521- 3526.

29. Morris, C.E., Klement, J.F., and McAllister, W.T. (1986). Cloning and expression of the bacteriophage T3 RNA polymerase gene. Gene, 41:193-200.

30. Leibowitz, M.J., McAllister, W.T. and Strohl, W.A. (1986). Viruses as causes of human cancer. J. of Med. Soc. of N.J., 83:603-608.

31. Klement, J.F., Ling, M.-L. and McAllister, W.T. (1986). Sequencing of DNA using T3 RNA polymerase and chain-terminating ribonucleoside triphosphate analogs. Gene Anal. Techn. 3:59-66.

32. Brown, J.E., Bailey, J.N., McAllister, W.T. (1986). Sequences of a region near the left end of bacteriophage T3 DNA that contains three promoters for the E. coli RNA polymerase. Nuc. Acids. Res. 14:4696-4696

33. Schaffner, A.R., Jorgensen, E.D., McAllister, W.T., and Hartmann, G.R. (1987). Specific labelling of the active site of T7 RNA polymerase. Nuc. Acids Res. 15:8773-8781.

34. Ling, M-L., Risman, S.S., Klement, J.F., McGraw, N., and McAllister, W.T. (1989). Abortive initiation by bacteriophage T3 and T7 RNA polymerases under conditions of limiting substrate. Nuc. Acids Res. 17:1605-1618.

35. Giordano, T.J., Deuschle, U., Bujard, H., and McAllister, W.T. (1989). Regulation of coliphage T3 and T7 RNA polymerases by the lac repressor-operator system. Gene, 84:209-219.

36. Deuschle, U., Pepperkok, R., Wang, F., Giordano, T.J., McAllister, W.T., Ansorge, W., and Bujard, H. (1989). Regulated expression of foreign genes in mammalian cells under the control of the bacteriophage T3 RNA polymerase and lac repressor. Proc. Nat. Acad. Sci. 86:5400-5404.

37. Giordano, T.J. and McAllister, W.T. (1990). Optimization of the hygromycin B resistance gene as a dominant selectable marker in mammalian cells. Gene, 88:285- 288.

38. Klement, J.F., Moorefield, M.B., Brown, J.E., Risman, S., and McAllister, W.T. (1990). Discrimination between T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three basepair region located 10-12 basepairs upstream from the start site. J. Mol. Biol. 215:21-29.

39. Joho, K.E., Gross, L.B., and McAllister, W.T. (1990). Identification of a region of the bacteriophage T3 and T7 RNA polymerases that determines promoter specificity. J. Mol. Biol. 215:31-39.

40. Sousa, R., Chung, Y.J., McAllister, W.T., Wang, B.C. and Lafer, E.M. (1990). Single crystals of a chimeric T7/T3 RNA polymerase with T3 promoter specificity and a non-processive T7 RNAP mutant. J. Biol. Chem. 265:21430-21432.

41. Rodriguez, D., Zhou, Y.W., Rodrigurz, J.R., Durbin, R.K., Jimenez, V., McAllister, W.T., and Esteban, M. (1990). Regulated expression of nuclear genes by T3 RNA polymerase and lac repressor, usiing recombinant vaccinia virus vectors. J. Virol. 64:4851-4857.

42. Zhou, Y., Giordano, T.J., Durbin, R.K., and McAllister, W.T. (1990). Synthesis of functional mRNA in mammalian cells by bacteriophage T3 RNA polymerase. Mol. Cell. Biol. 10:4529-4537.

43. Jorgensen, E.J., Durbin, R.K., and McAllister, W.T. (1991). Specific contacts between the phage T3, T7 and SP6 RNA polymerases and their promoters. J. Biol. Chem. 266:645-651.

44. Gross, L.B., Chen, W-J. and McAllister, W.T. (1992). Characterization of bacteriophage T7 RNA polymerase by linker insertion mutagenesis. J. Mol. Biol. 228:488-505.

45. Raskin, C.A., Diaz, G., Joho, K. and McAllister, W.T. (1992). Substitution of a single T3 residue into T7 RNA polymerase at position 748 results in a switch in promoter specificity. J. Mol. Biol. 228:506-515.

46. Luhrs, C.A., Raskin, C.A., Durbin, R., Wu, B., Sadasivan, E., McAllister, W.T., and Rothenberg, S.P. (1992). Transfection of a glycosylated phosphatidylinositol-anchored folate-binding protein complementary DNA provides cells with the ability to survive in low folate medium. J. Clin. Inv. 90:840-847.

47. Chen, W.J., Gross, L.B., Joho, K.E., and McAllister, W.T. (1992). A modified kanamycin-resistance cassette to facilitate two-codon insertion mutagenesis. Gene 111:143-144.

48. Diaz, G.A., Raskin,C.A. and McAllister, W.T. (1993). Hierarchy of base-pair preference in the binding domain of the bacteriophage T7 promoter. J. Mol. Biol., 229:805-811.

49. Raskin, C.A., Diaz, G.A. and McAllister, W.T. (1993). T7 RNA polymerase mutants with altered promoter specificities. Proc. Natl. Acad. Sci. USA, 90:3147-3151.

50. Macdonald, L., Zhou, Y. and McAllister, W.T. (1993). Termination and slippage by bacteriophage T7 RNA polymerase. J. Mol. Biol., 232:1030-1047.

51. McAllister, W.T., and Raskin, C.A. (1993). The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. Mol. Microbiol., 10:1-6.

52. McAllister, W.T. (1993). Structure and function of the bacteriophage T7 RNA polymerase (or, the virtues of simplicity). Cell. Mol. Biol. Res., 39:385-391.

53. Macdonald, L.E., Durbin, R.K., Dunn, J.J., and McAllister, W.T. (1994). Characterization of two distinct types of termination signals for the T7 RNA polymerase. J. Mol. Biol., 238:145-158.

54. He, B., McAllister, W.T. and Durbin, R.K. (1995). Phage RNA polymerase expression vectors that allow efficient expression of cloned genes in both prokaryotic and eukaryotic cells. Gene, 164:75-79.

55. Diaz, G.A., McAllister, W.T. and Durbin, R.K. (1996) The stability of abortively cycling T7 RNA polymerase complexes is dependent upon template conformation. Biochemistry 35:10837-10843.

56. He, B., Rong, M., Durbin, R.K., and McAllister, W.T. (1997) A mutant T7 RNA polymerase that is defective in RNA binding and blocked in the early stages of transcription. *J. Mol. Biol.* 265:276-288.

57. He, B., Rong, M., Lyakhov, D.L., Gartenstein, H., Diaz, G.A., Castagna, R.C., McAllister, W.T. & Durbin, R.K. (1997). Rapid generation and purification of mutant phage RNA polymerases. *Protein Expression & Purification* 9,142-151.

58. Lyakhov, D.L., He, B., Zhang, X., Studier, F.W., Dunn, J.J. & McAllister, W.T. (1997). Mutant T7 RNA polymerases with altered termination properties. *J.Mol.Biol.* **269**,28-40.

59. McAllister WT.(1997) Transcription by T7 RNA polymerase. *In*: Nucleic Acids and Molecular Biology. **11**,15-25, Eckstein F and Lilley D, eds, Springer-Verlag, Berlin.

60. Rong, M., He, B., McAllister, W.T. & Durbin, R.K. (1998). Promoter specificity determinants of T7 RNA polymerase. *Proc.Nat.Acad.Sci.U.S.A.* **95**: 515-519.

61. Rong, M., Durbin, R.K. & McAllister, W.T. (1998). Template strand switching by T7 RNA polymerase. *J.Biol.Chem.* **273**: 10253-10260.

62. Lyakhov, D.L., He, B., Zhang, X., Studier, F.W., Dunn, J.J. & McAllister, W.T. (1998). Pausing and termination by bacteriophage T7 RNA polymerase. *J.Mol.Biol.* **280**: 201-213.

63. He, B., Kukarin, A., Temiakov, D., Chin-Bow, S.T., Lyakhov, D.L., Rong, M., Durbin, R.K. & McAllister, W.T. (1998). Characterization of an unusual, sequence-specific termination signal for T7 RNA polymerase. *J. Biol. Chem.* **273**,18802-18811.

64. Gopal V, Brieba LG, Guajardo R, McAllister WT, Sousa R. (1999). Characterization of structural features important for T7 RNAP elongation complex stability reveals competing complex conformations and a role for the non-template strand in RNA displacement. *J. Mol. Biol.* **290**,411-431.

65. Place C, Oddos J, Buc H, McAllister WT, Buckle M. (1999). Studies of contacts between T7 RNA polymerase and its promoter reveal features in common with multisubunit RNA polymerases. *Biochemistry* **38**:4948-4957

66. Rong M, Castagna RC, McAllister WT. (1999). Cloning and purification of bacteriophage K11 RNA polymerase. *Biotechniques* **27**, 692-693.

67. Imburgio D, Rong M, Ma K, McAllister WT. (2000). Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **39**: 10419-10430.

68. Mentesana,P.E., Chin-Bow, S.T., Sousa, R., and McAllister, W.T. (2000). Characterization of halted T7 RNA polymerase elongation complexes reveals multiple factors that contribute to stability. J. Mol. Biol. **302**:1049-1062.

69. Temiakov, D., Mentesana, P.E., Ma, K., Mustaev, A., Borukhov, S. McAllister, W.T. (2001). The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc. Nat. Acad. Sci. (USA)*, **97**: 14109-14114. See Commentary by K. Severinov, *Proc. Nat. Acad. Sci. (USA)*, **98**: 5-7

70. Jiang M, Rong, M, Martin CT, McAllister WT. (2001). Interrupting the template strand of the T7 promoter facilitates translocation of the DNA during initiation, reducing transcript slippage and the release of abortive products. *J. Mol. Biol.* **310**: 509-522.

71. Imburgio, D. Anikin, M., McAllister, W.T. (2002). Effects of substitutions in a conserved DX_2GR motif found in many DNA-dependent nucleotide polymerases on transcription by T7 RNA polymerase. *J. Mol. Biol.* **319**: 37-51

72. Ma,K.; Temiakov,D.; Jiang,M.; Anikin,M.; McAllister,W.T. (2002). Major conformational changes occur during the transition from an initiation complex to an elongation complex by T7 RNA polymerase *J. Biol. Chem.* **277**:43206-43215.

73. Temiakov,D.; Anikin,M.; McAllister,W.T. (2002). Characterization of T7 RNA polymerase transcription complexes assembled on nucleic acid scaffolds. *J. Biol. Chem.* **277**:47035-47043.

74. Tahirov,T.; Temiakov,D.; Anikin,M.; Patlan,V.; McAllister,W.T.; Vassylyev,D.G.; Yokoyama,S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9Å resolution. *Nature* **420**:43-50.

75. Temiakov,D.; Tahirov,T.; Anikin,M.; McAllister,W.T.; Vassylyev,D.G.; Yokoyama,S. (2002). Crystallization and preliminary crystallographic analysis of T7 RNApolymerase elongation complex assembled on an RNA:DNA scaffold. *Acta Crystallographica* **59**:185-187

76. Kukarin, A., Rong, M.R., McAllister, W.T. (2002). Exposure of T7 RNA polymerase to the double stranded binding region of the promoter activates the enzyme to transcribe a single stranded template, *J. Biol. Chem.*, published online Novermber 18, 2002

Chapters in Books:

1. Bautz, E.K.F., McAllister, W.T., Kupper, H., Beck, E., and Bautz, F.A. (1974). Initiation of transcription by RNA polymerases of E. coli and phage T3. In "Control of Transcription", Biswal, B.B., Manadal, R.K., Stevens, A., and Conn, W.E., eds. Plenum Press, p. 115-123.

2. McAllister, W.T. and Sanders, M. (1980). Gene Families and Their Expression. In "Gene Families of Collagen and Other Proteins". D. Prockop, and P. Champe, eds. Elsevier, North Holland, Inc., pp. 179-190.

3. McAllister, W.T., Horn, N.J., Bailey, J.N., MacWright, R.S., Jolliffe, L., Gocke, C., Klement, J.F., Dembinski, D.R., and Cleaves, G.R. (1983). Specificity of the bacteriophage T3 and T7 RNA polymerases. In Gene Expression, UCLA Symposia on Molecular and Cellular Biology, New Series Vol. III, eds. D. Hamer and M. Rosenberg, Alan R. Liss, Inc., New York.

4. Sarver, N.S., Mitrani-Rosenbaum, S., Law, M.-F., McAllister, W.T., Byrne, J.C., and Howley, P.H. (1983). Bovine papillomavirus shuttle vectors. In Genetic Engineering: Principles and Methods, Vol. 5, pp. 173-190, ed., J.K. Setlow, Plenum Press, New York.

5. McAllister, W.T., McGraw, N.J., Morris, C.E. and Klement, J.F. (1985). Comparison of the bacteriophage T3 and T7 RNA polymerases. In Sequence Specificity in Transcription and Translation. UCLA Symposia on Molecular and Cellular Biology, New Series, Vol. 30, eds. R. Calendar and L. Gold, Alan R. Liss, Inc., New York, NY.

6. Morris, C.E., McGraw, N.J., Joho, K., Brown, J.E., Klement, J.F., Ling, M.L., and W.T. McAllister. (1987). Mechanisms of promoter recognition by the bacteriophage T3 and T7 RNA polymerases. In RNA Polymerase and Regulation of Transcription, eds., W.S. Reznikoff, R.R. Burgess, J.E. Dahlberg, C.A. Gross, M.T. Record, and M.P. Wickens, Elsevier, NY, pp. 47-58.

7. Jorgensen, E.D., Joho, K., Risman, S., Moorefield, M.B. and McAllister, W.T. (1989). Promoter recognition by bacteriophage T3 and T7 RNA polymerases. In DNA- Protein Interactions in Transcription. UCLA Symposia on Molecular and Cellular Biology, Alan R. Liss, Inc., New York, NY, pp.79-88.

8. McAllister, W.T. DNA chips and their impact on biotechnology and medicine. In Mediterranean Perspectives, Dowling College Press, Oakdale, NY (submitted)

9. Temiakov D, Karasavas PE, McAllister WT. (2000) Characterization of T7 RNA polymerase protein:DNA interactions during the initiation and elongation phases. Travers, A. A. and Buckle, M. Oxford, Oxford University Press. Protein:DNA interactions: A practical approach. Hames, B. D.

PATENT APPLICATIONS:

Issued:

1. Highly efficient dual T7/T3 promoter vector pJKF16 and dual SP6/T3 promoter vector pJFK15. US Patent No: 5,017,488 (May 21, 1991) William T. McAllister, John F. Klement.

2. Plasmid for the overproduction of bacteriophage T3 RNA polymerase, transcription vectors that carry a promoter recognized by its polymerase, gene coding for T3 RNA polymerase and application of these plasmids. US Patent No:5,037,045 (August 6, 1991), William T. McAllister.

3. Gene coding for a protein having T3 polymerase activity. US Patent No: 5,102,802 (April 7, 1992) William T. McAllister.

4. Method of producing a gene cassette coding for polypeptides with repeating amino acid sequences. US Patent No: 57089,406 (February 8, 1992) Jon L. Williams, Anthony J. Salerno, Ina Goldberg, and William T. McAllister.

5. Chimeric oligonucleotides and their use in the production of transcripts from nucleic acids. European Patent No. EP 0721988 (7/17/96) , Guillou-Bonnici, F, Levasseur, P., Cleuziat, P., McAllister, W.T., and Mallet, F., (bioMerieux, S.A.)

6. RNA-dependent RNA polymerase functioning preferably on a RNA matrix and promoter-dependent transcription process with said RNA-dependent RNA polymerase. US patent PCT/FR98/00635 (3/27/98). Cheynet-Sauvon, V., Arnaud-Barbe, N., Oriol, G., McAllister, W.T., Mandrand, B., Mallet, F. (bioMerieux, S.A.)

Disclosures:

1. Cloning and expression of the bacteriophage SP6 RNA polymerase gene and vectors for its use. (June 23, 1988) William T. McAllister, Russell Durbin, Steven Risman.

2. Expression of foreign genes in mammalian cells under the control of the bacteriophage T3 RNA polymerase. (January 3, 1989) William T. McAllister, Youwen Zhou.

3. Regulated transcription system based upon the bacteriophage T3 RNA polymerase and lactose repressor. (January 6, 1989) William T. McAllister, Thomas Giordano.

4. Development of a stable mouse cell line that expresses bacteriophage T3 RNA polymerase in the cytoplasm. (January 12, 1989) William T. McAllister, Thomas Giordano, Youwen Zhou.

5. Multipurpose cloning vector for expression of foreign genes in prokaryotic and eukaryotic cells under the control of the bacteriophage T3 RNA polymerase. (May 31, 1989) William T. McAllister, Russell K. Durbin.

6. T7 RNA polymerase mutants with altered promoter specificities. (May 4, 1992) William T. McAllister, Curtis A. Raskin.

7. Phage RNA polymerase expression vectors that allow efficient expression of cloned genes in both prokaryotic and eukaryotic cells. (February 2, 1995) Biao He, Russell K. Durbin, William T. McAllister

8. Improved T7 and T3 RNA polymerases having greater processivity. (April 14, 1993) Lynn MacDonald, John J. Dunn, William T. McAllister.

9. Phage RNA polymerases with relaxed specificity. (November 16, 1993) William T. McAllister, Russell K. Durbin, Curtis A. Raskin, George Diaz, Frederic R. Bloom, Jhy-Jhu Lin.

REFERENCES:

Available upon request

Cellular and Mo... ...r Biology Research, Vol. 39, pp. 385–391, 1993
Printed in the USA.

0968-8773/93 $6.00 + .00

Original Contribution

STRUCTURE AND FUNCTION OF THE BACTERIOPHAGE T7 RNA POLYMERASE (OR, THE VIRTUES OF SIMPLICITY)

WILLIAM T. MCALLISTER

Department of Microbiology and Immunology, Morse Institute of Molecular Genetics, SUNY Health Science Center, Box 44, 450 Clarkson Ave., Brooklyn, NY 11203-2098, USA

(*Received* 24 *August* 1993; *Accepted* 22 *September* 1993)

Abstract—A consideration of the properties of a number of mutants of T7 RNA polymerase, together with emerging structural information (Sousa et al., 1993) allows an interpretation of the the mechanics of transcription by this relatively simple RNA polymerase. Evidence indicating features in common with other nucleotide polymerases (such as DNA polymerases and reverse transcriptases) is reviewed.

Keywords—DNA polymerase, Reverse transcriptase, Promoter structure

INTRODUCTION

Unlike the multisubunit DNA-dependent RNA polymerases (RNAPs) of eukaryotic cells and bacteria, the RNAPs that are encoded by bacteriophage T7 and its relatives consist of a single species of protein that is capable of accurate transcription in the absence of any apparent need for auxiliary transcription factors (Chamberlin and Ryan, 1983). The striking simplicity of this transcription system makes it ideally suited for studies of RNA polymerase structure and function. The gene that encodes the phage RNAP has been cloned and may be overexpressed in bacterial cells, allowing genetic and biochemical manipulation of the enzyme (Davanloo et al. 1984). Importantly, T7 RNAP has now been crystallized, and a number of mutants that are altered in the transcription cycle have been characterized (Bonner et al., 1992; Patra et al., 1992; Gross et al., 1993; Sousa et al., 1993).

In our work, we have taken the approach of isolating or engineering T7 RNAP mutants with defined biochemical defects and asking whether these defects can be correlated with structural information so as to interpret the mechanism of transcription. We have identified important functional domains in the RNAP, and have found that the phage RNAP exhibits interesting structural and functional homologies to other simple nucleotide polymerases, such as DNA polymerases and reverse transcriptases. Although no extended sequence homologies exist between the phage RNAPs and the multisubunit RNAPs, there are intriguing clues that suggest a relationship between the phage enzymes and certain subunits of the more complex RNAPs. Studies of this class of RNAP will, therefore, contribute significantly to our understanding of nucleotide polymerization.

MATERIALS AND METHODS

Transcription reactions

Mutant RNAP have been previously described (Gross et al., 1993); the designation *ins*xxx indicates a linker insertion mutation that lies within or immediately preceding codon xxx. All transcription reactions were carried out in a volume of 10 μl containing: 20 mM Tris-HCl (pH 7.9), 8 mM $MgCl_2$, 2 mM spermidine-HCl, 1 mM dithiothreitol, 0.5 mM each of ATP, GTP, CTP, and UTP (Pharmacia, Ultrapure), and 1 μl cell extract (Gross et al., 1993). The products were resolved by electrophoresis in 20% polyacrylamide gels followed by autoradiography (ibid).

RESULTS

Enzyme domains involved in promoter recognition

T7 RNAP is the prototype of a class of single-subunit DNA-dependent RNAPs that includes the

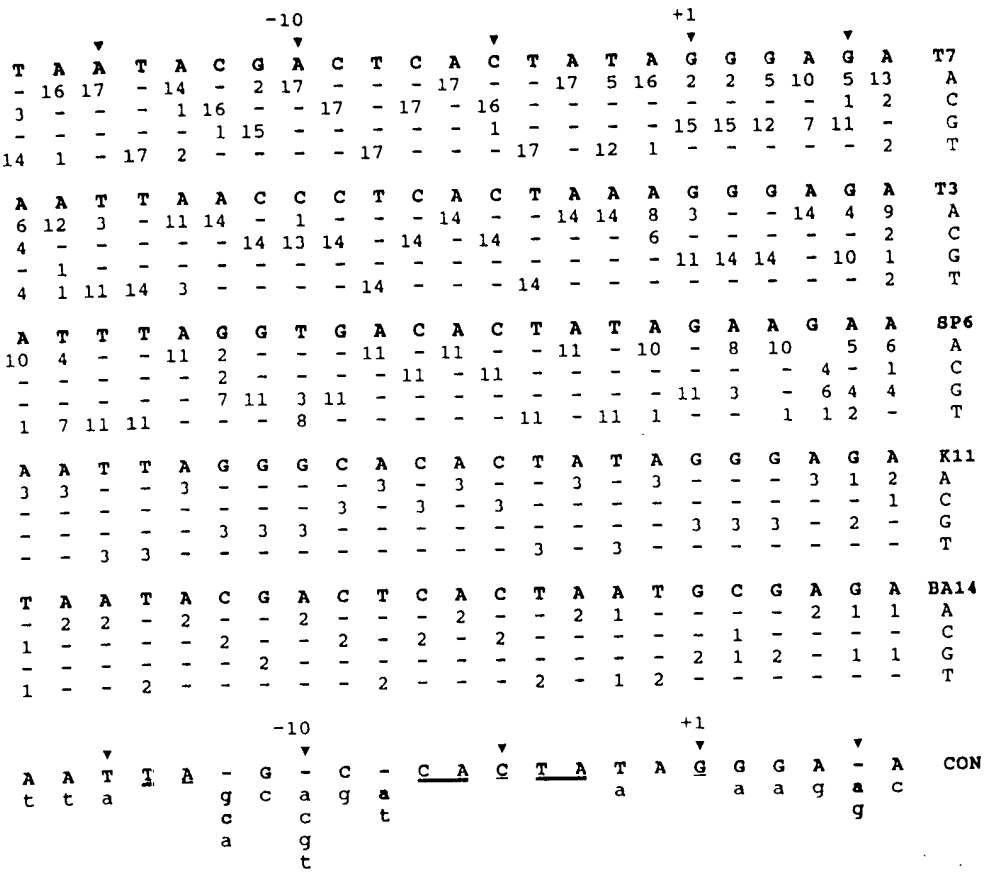


Fig. 1. Structure of phage promoters. Promoter sequences for T7, T3, K11, SP6, and BA14 RNAPs are compared. Below each consensus promoter is a summary of the frequency of each basepair in 17 natural T7 promoters, 14 T3 promoters, 11 SP6 promoters, 3 K11 promoters, and 2 BA14 promoters (Dunn and Studier, 1983; Brown et al., 1986; Beck et al., 1989; Dietz et al., 1990; Rush, personal communication; Hausmann, personal communication). The sequence of the nontemplate strand is shown; the start site of transcription is at +1. CON represents a common sequence of the five phage consensus promoters. Upper case, double underlined letters represent a bp that is conserved in all 47 promoters; upper case single underlined letters represent a bp that is conserved in all five consensus sequences; uppercase letters represent a conserved bp in at least three out of five consensus sequences; a dash (-) indicates a position that is not conserved. Lower case bold letters below the CON line indicate alternative bps that occur twice in the five consensus sequences; lower case letters indicate an alternative bp that occurs once.

RNAPs of related phages such as T3, SP6, and K11, as well as the mitochondrial RNAPs, and potentially, a chloroplast RNAP (for review, see McAllister and Raskin, 1993). Although the other phage RNAPs are closely related by sequence homology to T7 RNAP, each phage RNAP exhibits its own characteristic specificity. A comparison of the sequences of the phage promoters reveals a common 23 bp consensus sequence that extends from −17 to +6, with initiation at +1 (see Fig. 1).

A variety of biochemical and genetic experiments support the notion of two functional domains in the promoter—a binding domain that extends from −17 to −6, and an initiation domain that extends from −6 to +6 (see Fig. 2). All of the promoters share the same sequence from −6 to +1, indicating a common function for this region of the promoter. However, the sequences of the phage promoters differ significantly in the region from −9 to −12, suggesting that discrimination of specific promoter types may rely upon differences in this region. Mutations in the binding region have been observed to reduce the affinity of the polymerase for the promoter without having a great effect on the rate of initiation, whereas mutations in the initiation region have minor effects on the binding affinity but a greater effect on the rate of initiation (Chapman and Burgess, 1987; Chapman et al., 1988). Upon binding of the RNAP to the promoter, the DNA in the initiation region becomes melted open, as evidenced by a hypochromic shift and hypersensitivity of the nontemplate (NT) strand in this region to attack by single-strand specific endonucleases (Muller et al., 1989; Osterman and Coleman, 1981).

Base modification experiments indicate that a number of important contacts between the polymerase and the promoter are made within the major

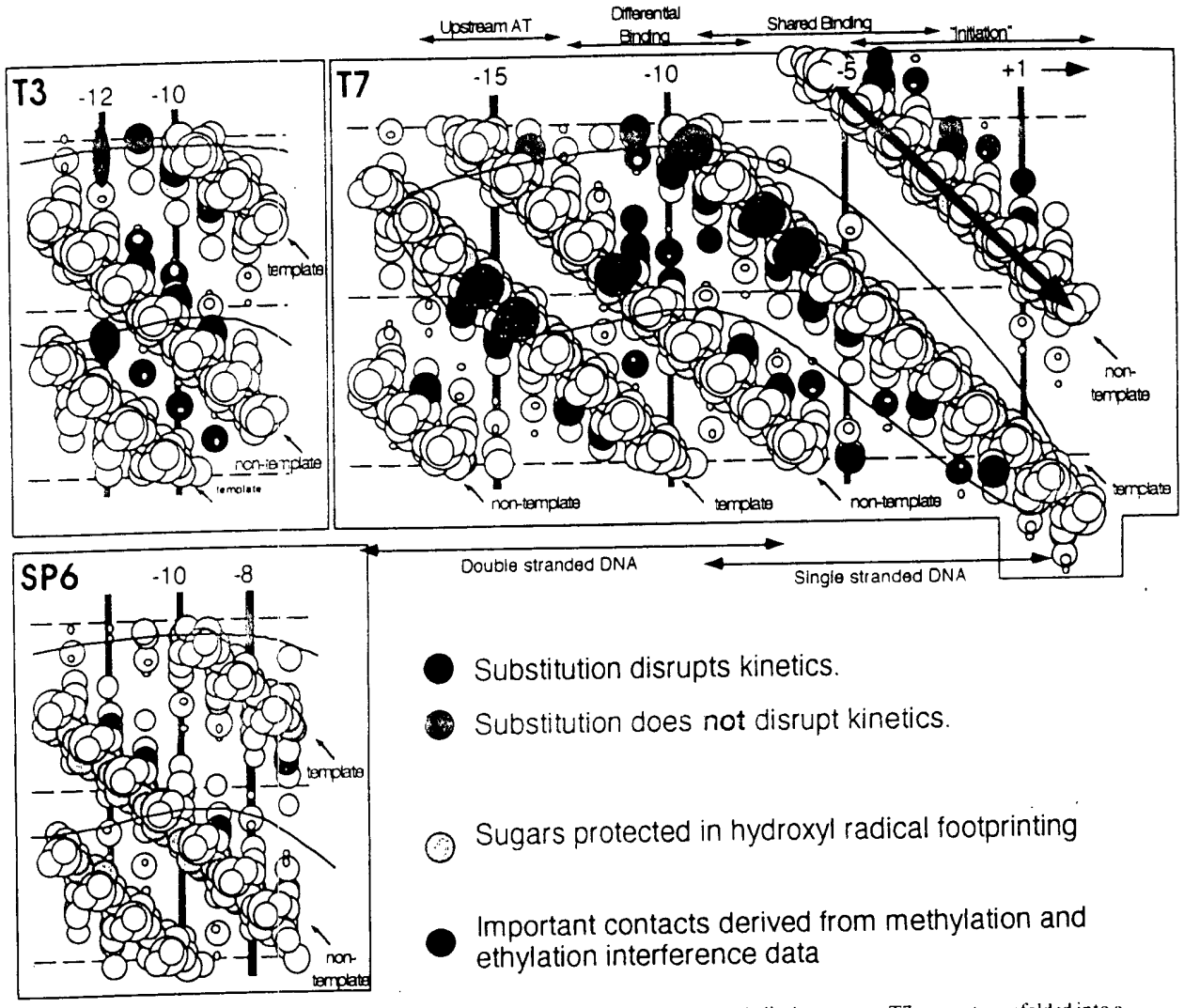


Fig. 2. Topography of RNA polymerase contacts. The drawing shows the double-helical consensus T7 promoter unfolded into a planar view; template, and nontemplate strands are indicated. Important structural elements in the promoter include: positions at which substitutions with modified bases affect the kinetics of initiation (Maslak et al., 1993; Schick and Martin, 1993), positions at which the sugar-phosphate backbone of the DNA is protected by polymerase binding as revealed by hydroxy-radical footprinting (Muller et al., 1989); and positions at which ethylation of the phosphate or methylation of the bases interferes with polymerase binding (Jorgensen et al., 1991). From these data, the contacts of the RNAP appear to involve major groove groups from −6 to −12, and minor groove contacts in the flanking regions on either side. Regions of the promoter that remain double stranded or are rendered partially single stranded during polymerase binding are indicated at the bottom (Osterman and Coleman, 1981; Muller et al., 1989). Other important regions such as the upstream AT-rich region, the region that is involved in promoter discrimination by individual RNAPs, the region in which the promoter sequences are highly conserved (shared binding), and the initiation region, are indicated at the top. Similar data for the T3 and SP6 promoters and their RNAPs are indicated in the side panels. Graphics were kindly provided by Dr. Craig Martin (University of Massachusetts).

groove and the flanking regions from bps −12 to −9, and it has been shown that the primary determinants of T3 vs. T7 promoter specificity are the bps at positions −11 and −10 (Jorgensen et al., 1991; Muller et al., 1989; Klement et al., 1990; Raskin et al., 1992). Substitution of these two bps in the T7 promoter with the corresponding bps found in the T3 promoter prevents recognition by T7 RNAP and simultaneously enables recognition by T3 RNAP (ibid).

To localize the region of the phage RNAP that is responsible for discrimination of these base pairs, hybrid T7/T3 RNAPs were constructed (Joho et al., 1990). In this way, the specificity determinant was localized to an 80 amino acid interval between residues 674 and 752. Within this interval the T7 and T3 RNAP amino acid sequences differ at only 11 positions. Site-directed mutagenesis of this region of the T7 RNAP indicates that a single amino acid is respon-

sible for discrimination of the −10 and −11 bps; when this residue (Asn) is substituted by the corresponding residue found in the T3 RNAP (Asp), the resulting mutant enzyme (T7-N748D) exhibits T3 promoter specificity, particularly for the bps found at −10 and −11 (Raskin et al., 1992). A consideration of the hierarchy of preference for each of the possible base pair combinations at −10 and −11 indicates that N748 makes direct contacts with bases on the nontemplate strand in a bidentate configuration (Diaz et al., 1993; Raskin et al., 1992). This interpretation is consistent with all of the genetic and biochemical data described above.

Substitution of other amino acids at position 748 has generated a collection of T7 RNAP mutants with altered specificities. Some of the mutant enzymes have specificities that correspond to those found in other phage RNAPs (e.g., the SP6 and K11 RNAPs), but others exhibit novel specificities not previously observed (Raskin et al., 1993). The location of residue N748 within the crystal structure of T7 RNAP is within a putative DNA binding cleft, at a position that would lie approximately one helical turn (35 Å) upstream from what is believed to be the active site (Sousa et al., 1993); see Fig. 3, and discussion below). This information serves to orient the RNA polymerase with respect to the promoter such that the direction of transcription along the template can be anticipated.

RNAP mutants blocked in other functions

To identify mutations that might affect other functions of the RNAP (catalysis, elongation, termination, etc.) we constructed 35 linker insertion mutants of T7 RNAP in which a 6 bp linker (two amino acids) was placed at various positions in the RNAP gene (Gross et al., 1993). These mutants were subjected to a variety of biochemical assays designed to detect blocks in key steps in the transcription cycle. A number of mutants with interesting biochemical phenotypes were identified, some of which are described below.

An additional region involved in promoter recognition

Among the linker insertion mutants were a class of RNAPs that retain nonspecific catalytic activity (i.e., they are able to synthesize poly rG on a poly dC template) but which have lost promoter-binding ability. Some of these mutations map near residue 748, as expected from the above discussion. However, other mutations map closer to the amino terminus of the protein. Two mutants in particular (*ins*144 and *ins*159, which consist of insertions within or before codons 144 and 159) are of particular interest because they lie near a region of T7 RNAP that exhibits significant sequence homology to region 2.4 of the bacterial sigma factor (Gross et al., 1993). This region of sigma factor is known to interact with base pairs in the −10 region of the *Escherichia coli* consensus promoter sequence (Helman and Chamberlin, 1988; Daniels et al., 1990; Siegele et al., 1989; Waldburger et al., 1990). In the crystal structure of T7 RNAP, this region is found within the DNA binding cleft, not far from the region defined by residue 748 (Fig. 3). Together, these two elements of the DNA binding cleft come in contact with the upstream region of the phage promoter, thus defining a sequence specific recognition element. The homology of this region to sigma factor suggests that additional common sequence elements may be found between the phage RNAPs and the multisubunit RNAPs.

Active site mutants

Another interesting class of mutants are those that retain promoter-binding activity, but have lost catalytic activity. Two interesting mutants within this class (*ins*640 and *ins*648) exhibit a characteristic defect in their ability to utilize double-stranded DNA templates but not single-stranded templates. For example, both of these enzymes exhibit significant activity on dC or dI-dC templates, but no activity on a dG:dC template (Gross et al., 1993). We reasoned that the defect in these enzymes might lie in their inability to melt open the double stranded helix, or failure to maintain an association with the template strand during elongation. This was confirmed by the use of synthetic promoters in which the promoter was "premelted" by virtue of the fact that the nontemplate strand in the initiation region was missing; whereas the wild-type enzyme is capable of initiating transcription from a fully duplex promoter as well as the premelted promoter, the two mutant enzymes were capable only of initiation from the premelted promoter (Gross et al., 1993).

The interpretation of these results with regard to the structure of T7 RNAP relied upon a potential similarity between the phage RNAPs and other nucleotide polymerases that was first observed by Delarue et al. (Delarue et al., 1990). These authors noted a homology in three sequence motifs (A–C) found in many nucleotide polymerases, including DNA polymerase and the single subunit DNA-dependent RNAPs. Two of these motifs (A and C) are also found in RNA-dependent RNA polymerases as well RNA-dependent DNA polymerases. In the structure of the Klenow fragment of *E. coli* DNA polymerase I (KF) these

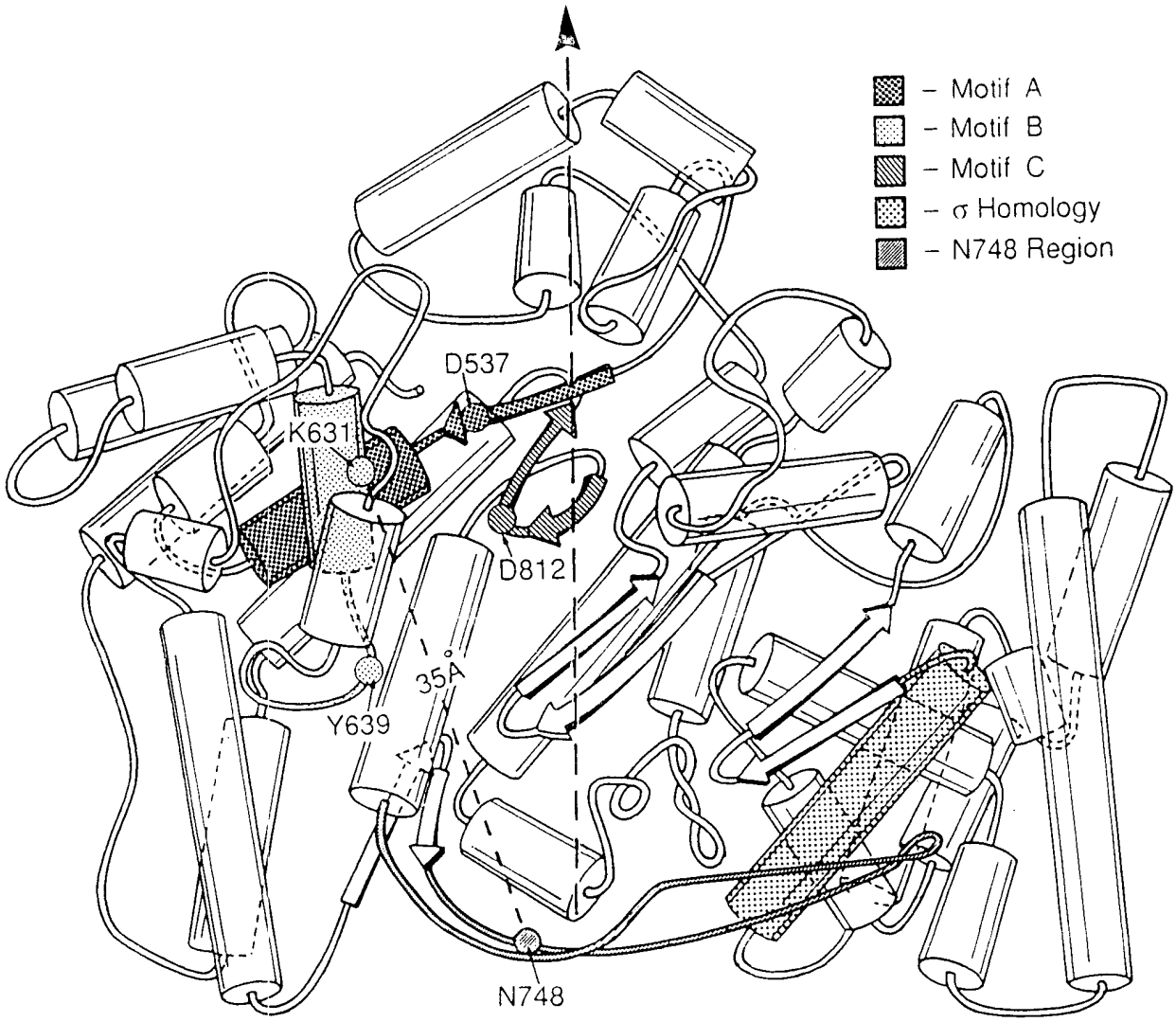


Fig. 3. Structure of T7 RNA polymerase. The schematic depicts T7 RNA polymerase looking into the DNA binding cleft; the axis of the cleft runs vertically, as indicated by the dashed arrow (adapted from Sousa et al., 1993). Structural motifs that are common to other nucleotide polymerases and which define the active site are indicated by selective shading (motifs A, B, and C), as is the region that exhibits homology with sigma factor region 2.4. Key catalytic residues are indicated. Residue N748, which is involved in contacts with the nontemplate bases at −10 and −11, lies on an extended loop at the base of the cleft. The distance from this residue to K631, which may be crosslinked to the initiating nucleotide, is 35 Å (approximately one turn of the double helix).

three regions are located near the active site (Ollis et al., 1985). This finding, and the observation that motif B differed in enzymes that utilize RNA vs. DNA as a template, led Delarue et al. to speculate that these polymerases may have evolved from a common precursor (or may use similar structural motifs to carry out common catalytic functions), and that motif B is likely to be involved in association with the template strand. The two mutations of interest in T7 RNAP (*ins*640 and *ins*648) lie within motif B. The inability of these mutant enzymes to melt open promoters or to remain stably associated with the template strand following initiation is consistent with the proposal that motif B is in association with the template strand.

Certain residues within motifs A, B, and C are highly conserved among all of the polymerases; these include, in particular, K631 in T7 RNAP, which lies in motif B. We and others have shown that this residue may be crosslinked to analogs of the initiating triphosphate, and that the crosslinked analog may subsequently serve as an acceptor in the formation of a phosphodiester bond with the next (incoming) nucleotide in a template-directed manner (Schaffner et al., 1987; Maksimova et al., 1991). Residue K631

must, therefore, be near the acceptor site in the initiation complex, consistent with its proximity to the template strand in the model described above.

More recent crystallographic data at higher resolution show a close structural correspondence between T7 RNA and KF, especially in the regions now referred to as the "polymerase-fold" (Sousa et al., 1993). A similar structural correspondence has been noted for the HIV reverse transcriptase, lending further support to the notion of a common catalytic mechanism for these enzymes (Kohlstaedt et al., 1992).

DISCUSSION

The convergence of genetic and biochemical approaches, as well as the availability of a high resolution crystal structure for T7 RNAP, make this a particularly exciting time to study the structure and function of an RNA polymerase. As a result of this and other work, considerable information is now available concerning the regions of the RNAP that are involved in promoter recognition, transcript elongation, and termination (for recent review, see (McAllister and Raskin, 1993). There is a growing body of evidence that supports the existence of a common polymerase fold among the simple nucleotide polymerases. This fold is likely to comprise the active site required for basic catalytic functions, and to contain elements that are involved in template binding and positioning of the active site. Other functions that are unique to the particular type of polymerase (e.g., promoter recognition and binding for the RNA polymerases, proofreading, and exonuclease functions for the DNA polymerases) are likely to be located elsewhere in the polymerase, possibly in auxiliary domains (see, for example, Fig. 3, in which the promoter recognition site is spatially quite separate from the putative active site).

What about the multisubunit RNA polymerases, do they also share homologies, or have they evolved along a different pathway? It is possible that as a result of the need to maximize the opportunity for regulation, multisubunit enzymes have distributed their corresponding functional motifs among multiple subunits. Sequence alignment programs may be unable to detect highly divergent motifs that are distributed among many protein subunits. A more fruitful approach may involve searching individual subunits for conserved motifs found in the phage-like RNA polymerases. The potential alignment between sigma factor and T7 RNAP suggests that this may prove to be an attractive method of analysis, although a functional role for this region of the phage RNAP must be confirmed. In any event, it is clear that studies of the structure and function of an elegantly simple RNAP like T7 will provide important clues for understanding the functioning of other polymerases.

Acknowledgements — This work was supported by NIH grant GM38147. I am grateful to the members of my laboratory for their performance of these studies, and to Rui Sousa for communicating results prior to publication.

REFERENCES

Beck, P.; Gonzalez, S.; Ward, C.; Molineux, I. Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9. J. Mol. Biol. 210:687–701, 1989.

Bonner, G.; Patra, D.; Lafer, E. M.; Sousa, R. Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. EMBO J. 11:3767–3775; 1992.

Brown, J. E.; Klement, J. F.; McAllister, W. T. Sequences of three promoters for the bacteriophage SP6 RNA polymerase. Nucl. Acids Res. 14:3521–3526; 1986.

Chamberlin, M. J.; Ryan, T. Bacteriophage DNA-dependent RNA polymerases. In: Boyer, P. D. The enzymes. New York: Academic Press; 1983:87–108.

Chapman, K.; Gunderson, S.; Anello, M.; Wells, R.; Burgess, R. Bacteriophage T7 late promoters with point mutations: Quantitative footprinting and in vivo expression. Nucl. Acids Res. 16:4511–4524; 1988.

Chapman, K. A.; Burgess, R. R. Construction of bacteriophage T7 late promoters with point mutations and characterization by in vitro transcription properties. Nucl. Acids Res. 15:5413–5432; 1987.

Daniels, D.; Zuber, P.; Losick, R. Two amino acids in an RNA polymerase sigma factor involved in the recognition of adjacent base pairs in the −10 region of a cognate promoter. Proc. Natl. Acad. Sci. USA 87:8075–8079; 1990.

Davanloo, P.; Rosenberg, A. H.; Dunn, J. J.; Studier, F. W. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 81:2035–2039; 1984.

Delarue, M.; Poch, O.; Tordo, N.; Moras, D.; Argos, P. An attempt to unify the structure of polymerases. Prot. Eng. 3:461–467; 1990.

Diaz, G. A.; Raskin, C. A.; McAllister, W. T. Hierarchy of base-pair preference in the binding domain of the bacteriophage T7 promoter. J. Mol. Biol. 229:805–811; 1993.

Dietz, A.; Weisser, H. J.; Kössel, H.; Hausmann, R. The gene for Klebsiella bacteriophage K11 RNA polymerase: Sequence and comparison with the homologous genes of phages T7, T3, and SP6. Mol. Gen. Genet. 221:283–286; 1990.

Dunn, J. J.; Studier, F. W. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. 166:477–535; 1983.

Gross, L.; Chen, W.; McAllister, W. T. Characterization of bacteriophage T7 RNA polymerase by linker insertion mutagenesis. J. Mol. Biol. 228:488–505; 1993.

Helman, J. D.; Chamberlin, M. J. Structure and function of bacterial sigma factors. Annu. Rev. Biochem. 57:839–872; 1988.

Joho, K. E.; Gross, L. B.; McGraw, N. J.; Raskin, C. A.; McAllister, W. T. Identification of a region of the bacteriophage T3 and T7 RNA polymerases that determines promoter specificity. J. Mol. Biol. 215:31–39; 1990.

Jorgensen, E. D.; Durbin, R. K.; Risman, S. S.; McAllister, W. T. Specific contacts between the bacteriophage T3, T7, and SP6 RNA polymerases and their promoters. J. Biol. Chem. 266:645–651; 1991.

Klement, J. F.; Moorefield, M. B.; Jorgensen, E.; Brown, J. E.; Risman, S.; McAllister, W. T. Discrimination between bacteriophage T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three base-pair region located 10 to 12

base-pairs upstream from the start site. J. Mol. Biol. 215:21–29; 1990.

Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science 256:1783–1790; 1992.

Maksimova, T. G.; Mustayev, A. A.; Zaychikov, E. F.; Lyakhov, D. L.; Tunitskaya, V. L.; Akbarov, A. K.; Luchin, S. V.; Rechinsky, V. O.; Chernov, B. K.; Kochetkov, S. N. Lys631 residue in the active site of the bacteriophage T7 RNA polymerase. Affinity labeling and site-directed mutagenesis. Eur. J. Biochem. 195:841–847; 1991.

Maslak, M.; Jaworski, M. D.; Martin, C. T. Tests of a model for promoter recognition by T7 RNA polymerase: Thymine methyl group contacts. Biochemistry 32:4270–4274; 1993.

McAllister, W. T.; Raskin, C. A. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. Mol. Microbiol. 10:1–6;1993.

Muller, D. K.; Martin, C. T.; Coleman, J. E. T7 RNA polymerase interacts with its promoter from one side of the DNA helix. Biochemistry 28:3306–3313; 1989.

Ollis, D. L.; Kline, C.; Steitz, T. A. Domain of *E. coli* DNA polymerase I showing sequence homology to T7 DNA polymerase. Nature 313:818–819; 1985.

Osterman, H. L.; Coleman, J. E. T7 ribonucleic acid polymerase-promotor interactions. Biochemistry 20:4884–4892; 1981.

Patra, D.; Lafer, E. M.; Sousa, R. Isolation and characterization of mutant bacteriophage T7 RNA polymerases. J. Mol. Biol. 224:307–318; 1992.

Raskin, C. A.; Diaz, G. A.; Joho, K. E.; McAllister, W. T. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. J. Mol. Biol. 228:506–515; 1992.

Raskin, C. A.; Diaz, G. A.; McAllister, W. T. T7 RNA polymerase mutants with altered promoter specificities. Proc. Natl. Acad. Sci. USA 90:3147–3151; 1993.

Schaffner, A.; Jorgensen, E. D.; McAllister, W. T.; Hartman, G. R. Specific labelling of the active site of T7 RNA polymerase. Nucl. Acids Res. 15:8773–8771; 1987.

Schick, C.; Martin, C. T. Identification of specific contacts in T3 RNA polymerase-promoter interactions: Kinetic analysis using small synthetic promoters. Biochemistry 32:4275–4280; 1993.

Siegele, D. A.; Hu, J. C.; Walter, W. A.; Gross, C. A. Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. J. Mol. Biol. 206:591–603; 1989.

Sousa, R.; Chung, Y. J.; Rose, J. P.; Wang, B. C. The crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. Nature 364:593–599; 1993.

Waldburger, C.; Gardella, T.; Wong, R.; Susskind, M. M. Changes in conserved region 2 of *Escherichia coli* sigma-70 affecting promoter recognition. J. Mol. Biol. 215:1–10; 1990.

Commentary

T7 RNA polymerase transcription complex: What you see is not what you get

Konstantin Severinov

Waksman Institute for Microbiology, Department of Genetics, 190 Frelinghuysen Road, Rutgers University, Piscataway, NJ 08854

Transcription, the first step of gene expression, is carried out by DNA-dependent RNA polymerases (RNAPs). RNAP catalyzes processive polymerization of RNA messages from NTP precursors by using one strand of DNA as a template. Although the reaction is similar to DNA polymerization (catalyzed by DNA polymerases), important differences exist. First, RNAPs are able to initiate synthesis of RNA from a nucleoside triphosphate, i.e., they do not require oligonucleotide primers. Second, the newly synthesized RNA chain is displaced normally from the DNA template. Thus, during transcription, the DNA strands are separated only in a short region around the catalytic center of the enzyme, the so-called transcription bubble. Because Watson–Crick interactions are required for template-dependent RNA synthesis, a transient RNA·DNA hybrid should exist inside the transcription bubble.

After promoter-complex formation, all RNAPs undergo abortive initiation–a catalytic synthesis of short, 2- to 8-nt-long, RNA oligomers that are rapidly synthesized and released from the complex. When the nascent RNA chain reaches a critical length of about 9 bases, it becomes stably associated with the transcription complex. RNAP then clears the promoter and elongates the nascent RNA chain in a fully processive manner. Despite its tight grip on nucleic acids, the elongating RNAP moves rapidly along the DNA and RNA chains until it encounters a termination signal, which typically consists of an RNA hairpin followed by a run of uridines. At such sites, RNAP transiently pauses, the stability of the elongation complex suddenly decreases, and the enzyme releases nucleic acids. The enzyme is now available to initiate transcription from promoters again.

The molecular determinants of transcription-complex stability and processivity are understood poorly. Several competing mechanistic models of RNAP function have been proposed in recent years. Much of the controversy centered around the length of RNA·DNA hybrid and its role (or lack thereof) in transcription. If the hybrid were relatively long, say, 8–9 base pairs, then the relative instability of the initial transcribing complex and the complex paused at a terminator could be explained by suboptimal hybrid length (refs. 1 and 2; Fig. 1). Conversely, the establishment of a full-length hybrid could explain the stabilization of the nascent RNA in transcription complex during promoter clearance. In contrast, if the hybrid remains short (less than 3 base pairs) throughout elongation, then complex stability should be determined primarily by the strength of the protein-nucleic acid interactions and/or conformational changes (refs. 2 and 3; Fig. 1). Establishment of the actual length of the hybrid might contribute also to our understanding of mechanisms of action of regulatory factors that modulate the rates of abortive initiation and promoter escape and the efficiency of transcription termination.

RNAPs seem to have arisen twice in evolution. A large family of multisubunit RNAPs includes bacterial enzymes, archeal enzymes, eukaryotic nuclear RNAPs, plastid-encoded chloroplast RNAPs, and RNAPs from some eukaryotic viruses. Members of this family exhibit extensive sequence and structural similarities (4, 5), suggesting that the mechanism of transcription is conserved highly within this group. The RNAP from *E. coli* (subunit composition of the catalytic core $\alpha^{I}\alpha^{II}\beta\beta'\omega$, molecular mass ≈380 kDa; it requires the specificity σ subunit to recognize and melt promoter DNA) is the best-studied enzyme of this family. An unrelated family of single-subunit RNAPs includes enzymes from bacteriophages and mitochondria as well as nuclear-encoded RNAPs of chloroplasts. Members of the latter family are related also to DNA polymerases and to reverse transcriptases (6). RNAP from bacteriophage T7 (M_r, ≈100 kDa, does not require additional factors for promoter recognition) is the best-studied member of this family.

Considerable evidence suggests that the hybrid length is ≈8 base pairs during transcription by *E. coli* RNAP. For example, RNA–DNA crosslinking experiments have established that only 8–9 bases closest to the 3′ end of the nascent RNA remain close to the template DNA strand in active elongation complex (7). Similarly, during initiation, the 5′ end of the nascent RNA remains close to DNA for about 8–9 bases and then branches away (8). Studies of the effects of base-analogue substitutions that either strengthen or weaken Watson–Crick interactions on elongation complex stability and termination efficiency also point to 8- to 9-bp hybrids (7). The upstream portion of the hybrid contributes the most to transcription complex stability (8, 9). Furthermore, a structural model of bacterial RNAP-elongation complex built by superimposing the positions of protein-nucleic acid crosslinks onto a high-resolution structure neatly accommodates an 8-bp hybrid (10).

Thus, if the nucleic acid scaffold plays an essential role in transcription, one would expect that the sizes of the transcription bubble and RNA·DNA hybrid would be similar in transcription complexes formed by enzymes of both classes.

Despite a lack of sequence similarity between the two RNAP families, the essential elements of the transcription cycle seem to be conserved (11). Thus, if the nucleic acid scaffold plays an essential role in transcription, one would expect that the sizes of the transcription bubble and RNA·DNA hybrid would be similar in transcription complexes formed by enzymes of both classes. It was surprising, therefore, when the structure of T7 RNAP

See companion article on page 14109 in issue 26 of volume 97.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.021535298. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.021535298

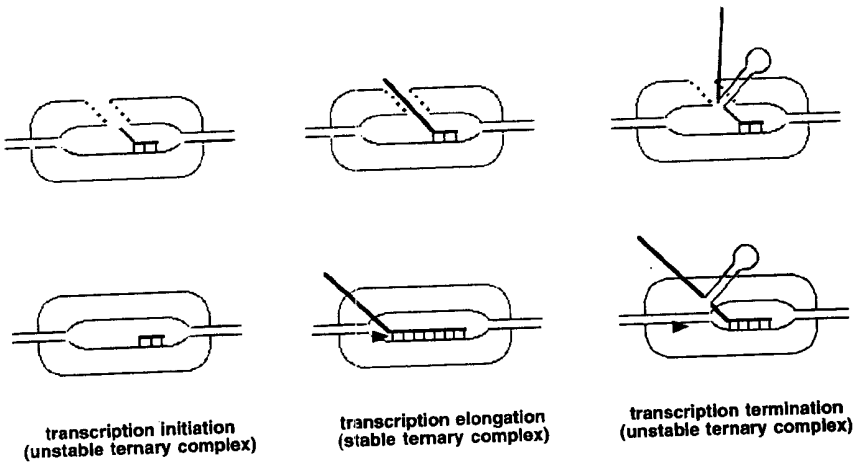


Fig. 1. Comparison of short (*Upper*) and long (*Lower*) RNA·DNA hybrid models of transcription complex. The stabilization of RNA (red) in the ternary transcription results from protein–RNA contacts (*Upper*) or is caused primarily by the establishment of full-length RNA·DNA hybrids (*Lower*). The crystal structure of transcribing T7 RNAP (12) looks like the schematic structure (*Upper Left*). Temiakov *et al.* (15) present data that the elongating T7 RNAP–transcription complex looks like the structure presented (*Bottom Center*) and, thus, is similar to the *Escherichia coli* RNAP (7). It is suggested that the structure (*Upper Center*) may correspond to transcription complex engaged in unproductive reiterative slippage synthesis.

complexed with a synthetic promoter and a trinucleotide RNA transcript revealed that only two RNA nucleotides closest to the catalytic center made Watson–crick interactions with the template strand of DNA (12). The 5′-proximal base of RNA appeared to peel away from the template, suggesting that in this system, the hybrid may be as short as 2–3 base pairs. Structural analysis suggested that (*i*) further extension of the hybrid would result in severe clashes with the T7 RNAP N-terminal domain, and (*ii*) a surface-exposed channel between the T7 RNAP thumb and the N-terminal domains is positioned appropriately to serve as an exit channel for the displaced RNA. If the difference between the RNA·DNA hybrid length in *E. coli* and T7 RNAP complexes is real, it has important mechanistic implications. For example, although both *E. coli* and T7 enzymes recognize identical terminators (13, 14), the mechanisms involved must be different (Fig. 1).

The ultimate way to resolve this impasse is to characterize structurally various intermediates of transcription cycle formed by both types of RNAPs. In the absence of such data, protein–RNA crosslinking and molecular modeling can provide useful information. In a recent issue of PNAS, Temiakov *et al.* (15) used RNA–DNA and RNA–protein crosslinking to study transcript elongation by T7 RNAP. Temiakov *et al.* incorporated a crosslinkable analogue of UMP, U*, into defined positions of the nascent RNA of artificially stalled, active T7 RNAP elongation complexes. The flexible spacer arm between the derivatized nucleotide base and the crosslinking group is long enough to allow the crosslinker to reach the complementary strand of a double-stranded nucleic acid. Thus, by following the appearance of RNA–DNA crosslinks between the derivatized nascent RNA and the template DNA strand, one can estimate the length of the RNA·DNA hybrid. The results are unambiguous and indicate that RNA–DNA crosslinks persist until the crosslinker is 9 base pairs upstream of the catalytic site (by convention, this position is referred to as −9). When the crosslinker is moved further away from the catalytic center, RNA–DNA crosslinks disappear, and RNA-protein crosslinks become prominent.

The RNA–DNA crosslinking experiment suggests a hybrid that is longer than observed in the structure of initiating T7 RNAP (12). To model the position of the hybrid in an elongating complex, Temiakov *et al.* (15) mapped the site of a crosslink between a short-range crosslinker incorporated into the nascent RNA 9 base pairs upstream of the catalytic center and RNAP. This critical position, where RNA is displaced from the hybrid, seems to contribute to elongation complex stability. By using a panel of chemical-mapping techniques in combination with RNAP mutants, the crosslink site was localized to within 7 T7 RNAP amino acids in the so-called specificity loop (16). The specificity loop is a phage RNAP-specific feature, which in the open promoter complex recognizes the double-stranded DNA 10–12 base pairs upstream of the transcription initiation start point. Strikingly, 2 amino acids within the 7-aa fragment that harbors the crosslink site are known to contact the promoter at positions −7, −10, and −11 and are critical in specific promoter recognition (16, 17). Thus, it seems that during promoter clearance, the contacts between the specificity loop and the upstream-promoter DNA are broken, and new contacts with RNA and possibly DNA at the upstream edge of the transcription bubble are established. The latter contacts may prevent collapse of the transcription bubble and thus stabilize RNA in the elongation complex. In addition, continued interactions between the specificity loop and DNA during elongation may contribute to sequence-specific pausing by T7 RNAP (18).

Localization of the position of ribonucleotide at the end of the hybrid allowed modeling of the overall position of the hybrid in the elongation complex (the 3′ end of the RNA in the hybrid is constrained, because the position of the catalytic center of the enzyme is known; ref. 16). The proposed trajectory results in few clashes, is consistent with much of the biochemical and genetic data, and is not consistent with the RNA-exit pathway suggested by the structure of the T7 RNAP initiation complex (12). In an independent study, Shen and Kang (19) mapped T7 RNAP–RNA crosslinks from either the derivatized 3′ end of the nascent RNA or from the −9 position of the nascent RNA in several stalled elongation complexes. Their results are in agreement with those of Temiakov *et al.* (15).

The picture of the T7 RNAP elongation complex that emerges from these studies is remarkably similar to our view of elongation complexes formed by multisubunit RNAPs. The RNA–DNA crosslinking result is superimposable with that obtained by Nudler *et al.* (7), who used the same experimental approach with *E. coli* RNAP. Thus, during elongation, the RNA·DNA hybrid appears to be 8–9 bp in length in both types of RNAP. In the T7 RNAP elongation complex, the specificity loop is proposed to act as a wedge that both separates the nascent RNA from the DNA template and possibly maintains the upstream edge of the transcription bubble. In the structural model of bacterial RNAP elongation complex (10), an evolutionarily conserved feature of the largest subunit, the so-called "rudder," seems to play an analogous role. Interestingly, the rudder may also be involved in promoter recognition, because it is close to a region of the specificity subunit σ that recognizes promoter positions −9/−12 (20). Site-directed deletion mutagenesis reveals that the rudder indeed contributes to bacterial RNAP elongation-complex stability by preventing the displacement of the nascent RNA by the nontemplate DNA strand (K. Kuznedelou, N. Korzheva, A. Mustaev, and K.S., unpublished observa-

tions). Unfortunately, similar experiments to demonstrate the role of specificity loop in T7 RNAP elongation and complex stability are complicated, because the specificity loop is required strictly for promoter recognition.

The inconsistency between the structure of the T7 RNAP initiation complex obtained by x-ray crystallography and the view that emerges from biochemical studies could be explained by large conformational changes that may occur during the transition from an unstable-initiation complex to a stable-elongation complex. A more likely explanation is that in the crystal structure, T7 RNAP had been captured in an act of unproductive synthesis. When the initial transcribed sequence codes for three Gs in RNA, T7 RNAP can synthesize long chains of poly(G) through repetitive cycles of slippage of the nascent RNA along the template strand and addition of the next GMP (21). During such reiterative synthesis, RNAP does not leave the promoter, and naturally the hybrid can be only 3 bp or less. The addition of the nucleoside triphosphate specified by the fourth position of the template (provided it does not code for a G) inhibits slippage and leads to productive initiation. The crystals of transcribing T7 RNAP complex were obtained in the presence of GTP and chain-terminating α,β-methylene-ATP. Thus, the expected RNA product should have been GGGA. However, the adenosine nucleotide is not present in the structure, and subsequent biochemical data indicate that the ATP analogue has no effect on the slippage reaction (22). Thus, the exit pathway suggested by the early peeling RNA could be that of a slipped transcript and may not be used during productive elongation.

Multisubunit RNAPs also undergo transcript slippage, and the strength of some promoters is regulated by variation of productive vs. unproductive, reiterative initiation events (23). As is the case with T7 RNAP, the slippage reaction in *E. coli* occurs at the end of a run of three or more identical base pairs in the initial transcribed sequence. Thus, during slippage, the RNA–DNA hybrid is short, suggesting that the RNA-exit pathway of the complex engaged in the slippage synthesis is different from that in the productive complex (10). Interestingly, in bacterial RNAP a surface-exposed channel between the two domains of the second largest subunit exists (4) that is positioned analogously to the putative RNA-exit channel seen on the T7 RNAP transcribing complex (12). Consistent with this idea, in bacterial RNAP, mutations that alter the position of one of the second largest subunit domains or change residues that lie between the two domains result in dramatic changes in the efficiency of reiterative synthesis (24, 25). Future comparative structural analysis of transcription intermediates, in conjunction with biochemical and genetic analyses, should determine the true extent of functional convergence that nature has come up with and solve an identical problem of transcribing a defined fragment of DNA with two different protein machines.

I thank Dr. Karen Adelman for her comments. This paper was supported by National Institutes of Health Grant RO1 59295.

1. Yager, T. D. & von Hippel, P. H. (1991) *Biochemistry* **30,** 1097–1118.
2. Nudler, E. (1999) *J. Mol. Biol.* **288,** 1–12.
3. Uptain, S. M., Kane, C. M. & Chamberlin, M. J. (1997) *Annu. Rev. Biochem.* **66,** 117–172.
4. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999) *Cell* **98,** 811–824.
5. Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R. & Kornberg, R. D. (2000) *Science* **288,** 640–649.
6. McAllister, W. T. & Raskin, C. A. (1993) *Mol. Microbiol.* **10,** 1–6.
7. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. (1997) *Cell* **89,** 33–41.
8. Korzheva, N., Mustaev, A., Nudler, E., Nikiforov, V. & Goldfarb, A. (1998) *Cold Spring Harbor Symp. Quant. Biol.* **63,** 337–345.
9. Kireeva, M. L., Komissarova, N., Waugh, D. S. & Kashlev, M. (2000) *J. Biol. Chem.* **275,** 6530–6536.
10. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. & Darst, S. A. (2000) *Science* **289,** 619–625.
11. McAllister, W. T. (1997) *Nuceic Acids Mol. Biol.* **11,** 15–24.
12. Cheetham, G. M. & Steitz, T. A. (1999) *Science* **286,** 2305–2309.
13. Christiansen, J. (1988) *Nucleic Acids Res.* **16,** 7457–7476.
14. Macdonald, L. E., Durbin, R. K., Dunn, J. J. & McAllister, W. T. (1994) *J. Mol. Biol.* **238,** 145–158.
15. Temiakov, D., Mentesana, P. E., Ma, K., Mustaev, A., Borukhov, S. & McAllister, W. T. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 14109–14114. (First Published November 28, 2000; 10.1073/pnas.250473197)
16. Cheetham, G. M., Jeruzalmi, D. & Steitz, T. A. (1999) *Nature (London)* **399,** 80–83.
17. Raskin, C. A., Diaz, G., Joho, K. & McAllister, W. T. (1992) *J. Mol. Biol.* **228,** 506–515.
18. He, B., Kukarin, A., Temiakov, D., Chin-Bow, S. T., Lyakhov, D., Rong, M., Durbin, R. K. & McAllister, W. T. (1998) *J. Biol. Chem.* **273,** 18802–18811.
19. Shen, H. & Kang, C. (2000) *J. Biol. Chem.*, in press.
20. Arthur, T. M. & Burgess, R. R. (1998) *J. Biol. Chem.* **273,** 31381–31387.
21. Martin, C. T., Muller, D. K. & Coleman, J. E. (1988) **27,** 3966–3974.
22. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. (2000) *Biochemistry* **29,** 10419–10430.
23. Liu, C., Heath, L. S. & Turnbough, C. L., Jr. (1994) *Genes Dev.* **8,** 2904–2912.
24. Martin, E., Sagitov, V., Burova, E., Nikiforov, V. & Goldfarb, A. (1992) *J. Biol. Chem.* **267,** 20175–20180.
25. Jin, D. J. (1996) *J. Biol. Chem.* **271,** 11659–11667.

Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution

Tahir H. Tahirov*†, Dmitry Temiakov†‡, Michael Anikin‡, Vsevolod Patlan§, William T. McAllister‡, Dmitry G. Vassylyev§‖ & Shigeyuki Yokoyama§‖¶#

** High Throughput Factory, § Structurome Research Group, and ‖ Cellular Signaling Laboratory, RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan*
‡ Morse Institute for Molecular Genetics, Department of Microbiology, SUNY Health Science Center, 450 Clarkson Avenue, Brooklyn, New York 11203, USA
¶ RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan
 # Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
† These authors contributed equally to this work

The single-subunit bacteriophage T7 RNA polymerase carries out the transcription cycle in an identical manner to that of bacterial and eukaryotic multisubunit enzymes. Here we report the crystal structure of a T7 RNA polymerase elongation complex, which shows that incorporation of an 8-base-pair RNA–DNA hybrid into the active site of the enzyme induces a marked rearrangement of the amino-terminal domain. This rearrangement involves alternative folding of about 130 residues and a marked reorientation (about 130° rotation) of a stable core subdomain, resulting in a structure that provides elements required for stable transcription elongation. A wide opening on the enzyme surface that is probably an RNA exit pathway is formed, and the RNA–DNA hybrid is completely buried in a newly formed, deep protein cavity. Binding of 10 base pairs of downstream DNA is stabilized mostly by long-distance electrostatic interactions. The structure implies plausible mechanisms for the various phases of the transcription cycle, and reveals important structural similarities with the multisubunit RNA polymerases.

Gene expression in all organisms requires messenger RNA synthesis by DNA-dependent RNA polymerase (RNAP). These enzymes can be divided into two classes: multisubunit (bacteria, archaea, eukaryotes) and single-subunit (some bacteriophages, mitochondria, chloroplasts) RNAPs. Although they share no apparent sequence or structural homology, the RNAPs of both classes carry out the basic steps of transcription in an identical manner[1]. To initiate RNA synthesis, the enzyme must bind to a specific promoter DNA sequence that lies upstream of the start site for transcription, separate (melt) the two strands of the DNA in the vicinity of the start site (forming a transcription bubble), and begin RNA synthesis using the coding strand of the downstream DNA as a template and a single ribonucleotide as a primer. During the early stages of transcription, contacts between the RNAP and the upstream promoter sequences are maintained while the active site translocates downstream, resulting in the formation of a short RNA–DNA hybrid and extension of the transcription bubble[2–5]. The T7 RNAP initiation complex (IC) is unstable and repeatedly releases short (3–8 nucleotides) abortive RNA products before it undergoes a transition to form a stable elongation complex (EC)[6]. The transition starts when the RNA–DNA hybrid reaches a length of 8–9 base pairs (bp) and results in promoter release, collapse of the melted promoter region, and displacement of the 5′ end of the nascent RNA[4,5,7]. During elongation the length of the RNA–DNA hybrid is maintained at 7–8 bp and the transcription bubble closes just after the RNA chain peels away from the DNA template[8–10].

The solution of several RNAP structures has resulted in a breakthrough in our understanding of the functional aspects of transcription[11–20]. The structures of T7 RNAP–nucleic acid complexes revealed that promoter binding involves three principal structural motifs of the RNAP: an (A + T)-rich recognition loop interacts with the −17 region; a specificity loop makes base-specific contacts around −9; and an intercalation loop involving Val 237 facilitates promoter melting and stabilizes the upstream edge of the separated transcription bubble between −5 and −4 (the start site for transcription is designated as +1). The structure of an early IC in which the first three bases of the template strand have been transcribed is essentially unchanged from that of the binary promoter complex, indicating that the initial stages of transcription may be achieved without major changes in enzyme structure. However, the structure of the IC did not allow space in the active site for an RNA–DNA hybrid greater than 3 bp, and therefore did not provide a plausible mechanism for further transcription progress. The crystal structure of the T7 RNAP elongation complex reported here reveals that incorporation of an 8-bp-long RNA–DNA hybrid into the active site results in unprecedented structural alterations of protein structure, thereby resolving the principal differences between previous biochemical and structural data[9,14,21]. Although the conformation of the carboxy-terminal portion of T7 RNAP (which includes the active site) resembles that of polymerase I-like DNA polymerases[11,22,23], the overall organization of the EC is remarkably similar to that of the multisubunit RNAPs, reflecting the common functions of these two classes of RNAPs.

Structure determination and overall structure

Elongation complexes were formed and crystallized as described[24,25]. The structure was refined at 2.9 Å resolution to a final $R = 23.5\%$ and $R_{free} = 28.4\%$ (Fig. 1a, b; see also Supplementary Table 1).

The RNAP in the EC has a shell-like, highly porous architecture (Fig. 1c–e). The downstream DNA is bound in a deep groove and enters through a wide passage to a cavity that contains an 8-bp RNA–DNA hybrid. The axis of the hybrid is nearly perpendicular to the entering DNA, as in the yeast RNA polymerase II elongation complex[17]. The structure contains partly accessible channels, which presumably correspond to binding sites for the template and non-template DNA strands, and prominent pores for entry of the substrate and exit of the RNA product. The positive charge covering

nature This advance online publication ('AOP') *Nature* paper should be cited as 'Author(s) *Nature* advance online publication, 9 October 2002 (doi:10.1038/nature01129)'. Once the print version (identical to the AOP) is published, the citation becomes 'Author(s) *Nature* volume, page (year)' advance online publication, 9 October 2002 (doi:10.1038/nature01129)'.

almost the entire interior of the molecule extends through the pores and channels to the external surface. The channels and pores are features characteristic only to the EC.

Protein structure

During the formation of the EC, the N-terminal domain of the RNAP (residues 2–266) undergoes remarkable structural reorganization as compared with the IC (Fig. 2). This involves a marked reorientation of a core subdomain (72–151, 206–257), which remains unaltered, and an alternative folding of approximately 130 residues. These changes result in the formation of three structural elements: an N-terminal extension (N-subdomain; residues 2–71), a central flap-like subdomain (152–205), and a C-terminal linker that connects the N-terminal domain to the C-terminal portion of the enzyme (C-linker; 258–266) (Fig. 2).

Relative to the IC (Protein Data Bank accession number 1QLN[14]; Fig. 2b, c), the core subdomain does not change its internal structure (root mean square deviation, r.m.s.d. = 0.7 Å over 524 main chain atoms) but moves as a rigid body (35 Å translation and 130° rotation) to its final position in the EC (Fig. 2a, c). This reorientation opens space in the active site to accommodate the expanding RNA–DNA hybrid. Of note, the core exhibits high intrinsic structural similarity (but lacks sequence homology) to a C-terminal region of T7 RNAP and to a lesser extent to the corresponding segment of pol I-like DNA polymerases (Taq DNAP; Protein Data Bank accession number 1TAQ[26]) (Fig. 3). This suggests that the core subdomain evolved from duplication of a conserved structural element.

In the IC (Fig. 2b, c), the N-subdomain does not interact with nucleic acids, and its fold consists of seven structural motifs: loop-helix-loop-helix-loop-helix-disordered region. In the EC (Fig. 2a, c), the N-subdomain adopts a more compact loop-helix-loop-helix-loop conformation that constitutes a portion of the binding site for the RNA–DNA hybrid. Despite the marked difference in folding, the N-subdomains in the IC and EC are located on the same side of the RNAP molecule, so that a short α-helical fragment (31–43) coincides in both structures. This suggests that the N-subdomain keeps a fixed orientation during the structural transition of the RNAP, and that all alterations in this domain result from alternative folding.

A long, unfolded loop (165–205) intervenes between the N- and C-terminal portions of the core in the IC (Fig. 2b, c). In the EC (Fig. 2a, c) this loop is folded into the flap subdomain, which consists of a helix-turn-helix (HTH) motif followed by a C-terminal loop. In the EC, the flap subdomain spans the interval between the upstream end of the RNA–DNA hybrid and downstream DNA.

Finally, whereas the short C-linker connecting the N-terminal domain to the C-terminal domain has a random coil conformation in the IC (Fig. 2b, c), it adopts an α-helical conformation in the EC (Fig. 2a, c) and extends the C-terminal α-helix of the core subdomain towards the end point of the N-terminal domain.

The C-terminal, DNA polymerase-like portion of the RNAP,

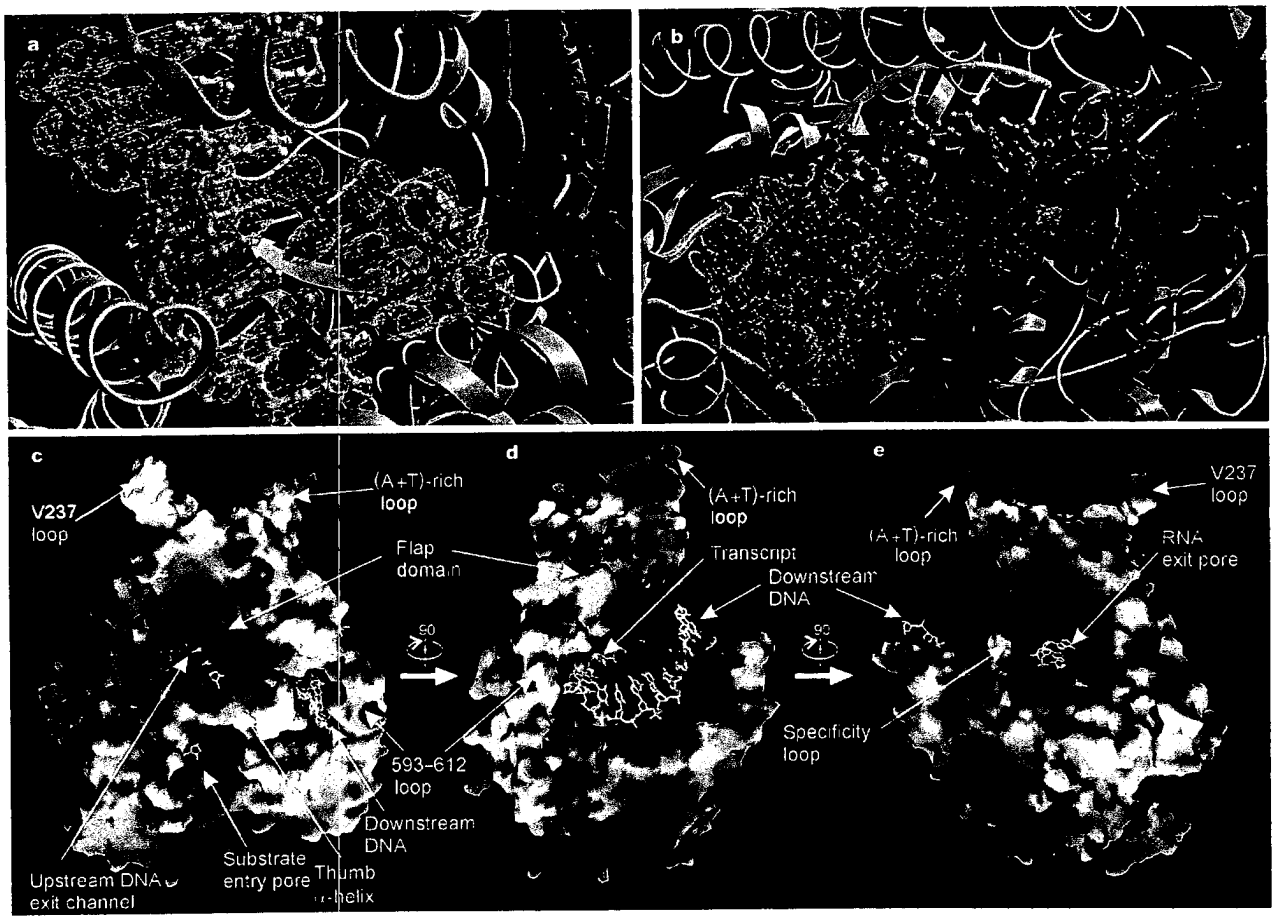


Figure 1 The T7 RNAP EC crystal structure. The RNA (light yellow), DNA template (red) and non-template (blue) strands are shown as a 'ball-and-stick' model. **a**, **b**, The $|F_o - F_c|$ omit electron density map (3.2σ contour level; green) produced for the RNA–DNA hybrid (**a**) and for the long N-terminal α-helix (residues 31–62, ball-and-stick model; magenta) (**b**) is superimposed on the atomic model. The rest of the protein is represented by a ribbon diagram. **c**–**e**, Three views of the EC. The protein surface is coloured according to the electrostatic potentials (positive, negative and neutral are dark blue, red and white, respectively).

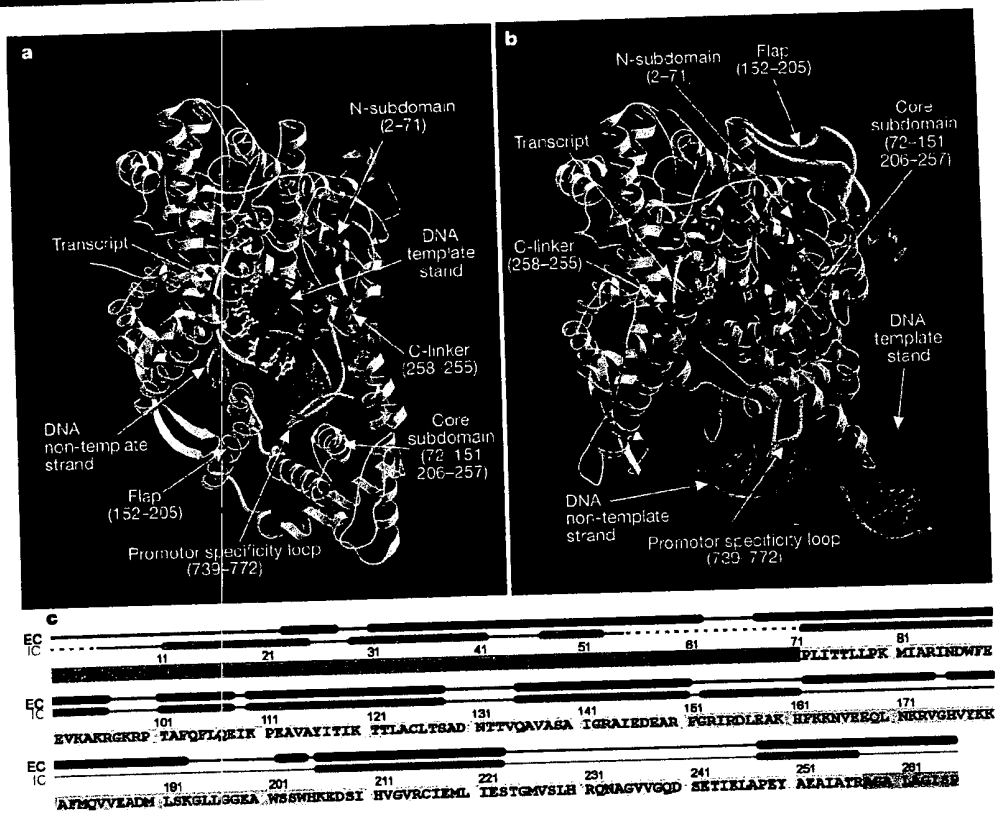


Figure 2 Comparison between the T7 RNAP EC and IC structures. **a**, **b**, The EC (**a**) and the early IC (**b**) ribbon diagrams are shown in the same orientation with respect to the unaltered C-terminal portion (white). The view in **a** corresponds to that of Fig. 1c, but rotated 180° around the horizontal axis. The nucleic acids are shown as ribbon (phosphate backbone) and ball-and-stick (bases) models. **c**, Amino acid sequence of the N-terminal domain; secondary structures for the EC and the early IC are shown above (cylinders, α-helices; thin lines, coils; dashed lines, disordered regions). The colour scheme is the same as in **a** and **b**.

which consists of three major domains designated as 'fingers', 'palm' and 'thumb'[11,12], remains largely unaltered in the EC as compared with the IC (Fig. 2a, b). The primary exception is the specificity loop (739–772)[13,27,28] (Fig. 2a, b). A substantial shortening of the loop hairpin (by about 13 Å) is probably the result of outlooping of the C-terminal strand in the vicinity of residues 757–767, as indicated by weak but visible electron density in this region. Furthermore, the tip of the loop is buried in a hydrophobic cavity formed by the flap and core subdomains, and the region of the loop that is used for base-specific promoter recognition in the IC is shifted. As a result of these changes, and reorientation of the core subdomain, the specificity loop becomes separated from the two elements in the core subdomain that were involved in promoter binding in the IC.

There are two other noticeable alterations in the C-terminal portion of the enzyme. First, the loop at the tip of the fingers domain (593–612) moves towards the downstream DNA to form part of the DNA-binding site (Fig. 1c, d). Second, the N-terminal end of the long α-helix in the thumb domain bends towards the

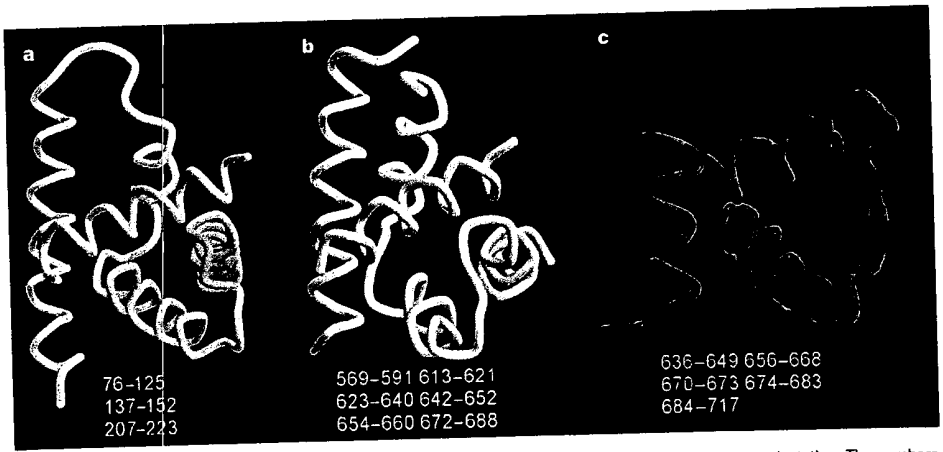


Figure 3 Structural homology of the T7 RNAP core subdomain. **a**–**c**, The T7 RNAP core (**a**), the C-terminal homologous region (**b**), and the homologous fragment of the Taq DNA polymerase (**c**) are shown in the same orientation. The numbers of residues used for the superimposition are shown at the bottom.

fingers domain to complete the formation of the pore that leads to the active site (Fig. 1c).

Protein–nucleic acid interactions

The position of the downstream DNA is determined largely by long-range electrostatic interactions (Fig. 4a) with parts of the fingers subdomain and with residues in the first turn of the flap subdomain. The DNA strands are separated just before the template strand enters the active site by interaction with the N-terminal part of a helix in the fingers subdomain (642–662), which we designate a 'downstream DNA zipper'. DNA melting involves stacking of the template base at +2 on the Phe 644 side chain of the zipper, which appears to cause the base at +1 to flip out, thereby directing the template strand towards the active site (Fig. 4b). (Nucleotide positions in the EC are designated relative to the active site at +1.) The first unpaired nucleotide of the non-template strand (at +1) is directed towards a deep groove formed by a loop (593–610) and a helix-turn-helix fragment (647–675) from the fingers subdomain, part of the flap subdomain, and the N-terminal portion of the long α-helix from the thumb subdomain (Fig. 1c, d). This groove extends to the template strand exit located between the second turn of the flap subdomain and the C-terminus of the long helix of the N-terminal extension (Fig. 1c).

The 8-bp RNA–DNA hybrid in the EC adopts an underwound A-form conformation of the double helix that is similar to the conformation of the RNA–DNA hybrid in the yeast RNA polymerase II EC (Fig. 1a; r.m.s.d. 0.6 Å)[17]. The hybrid-binding cavity complements this shape and has a surface that consists of alternating hydrophobic and positively charged residues, which may facilitate translocation (Fig. 4a). The first 3 bp at the downstream end of the hybrid interact with the protein in a similar fashion as in the T7 IC (Fig. 4a). However, the interactions of the 3 bp at the upstream end of the hybrid involve newly folded elements of the EC. Most notable are interactions of template strand nucleotides at positions −6, −7 and −8 with the C-terminal portion (50–60) of the long helix from the N-subdomain, and the RNA nucleotides at positions −6 and −7 with the flap subdomain (Fig. 4a).

The upstream boundary of the RNA–DNA hybrid cavity is defined by the C-terminal part of a loop-helix motif (62–71) that connects the long helix of the N-subdomain with the core subdomain. The Gln 58, Glu 63 and Asp 66 side chains in this loop project towards the plane of the base pair at the upstream end of the hybrid at a distance of about 7.3 Å (Fig. 4a, c), which may allow hybrid extension by 1 bp without serious protein alterations.

RNA exit and substrate entry pores

In the EC, a highly positive pore leads from the upstream end of the RNA–DNA hybrid to the surface, and probably provides an RNA exit pathway[29,30] (Fig. 1e). The pore is formed by part of the reorganized specificity loop, the newly formed C-linker, and part

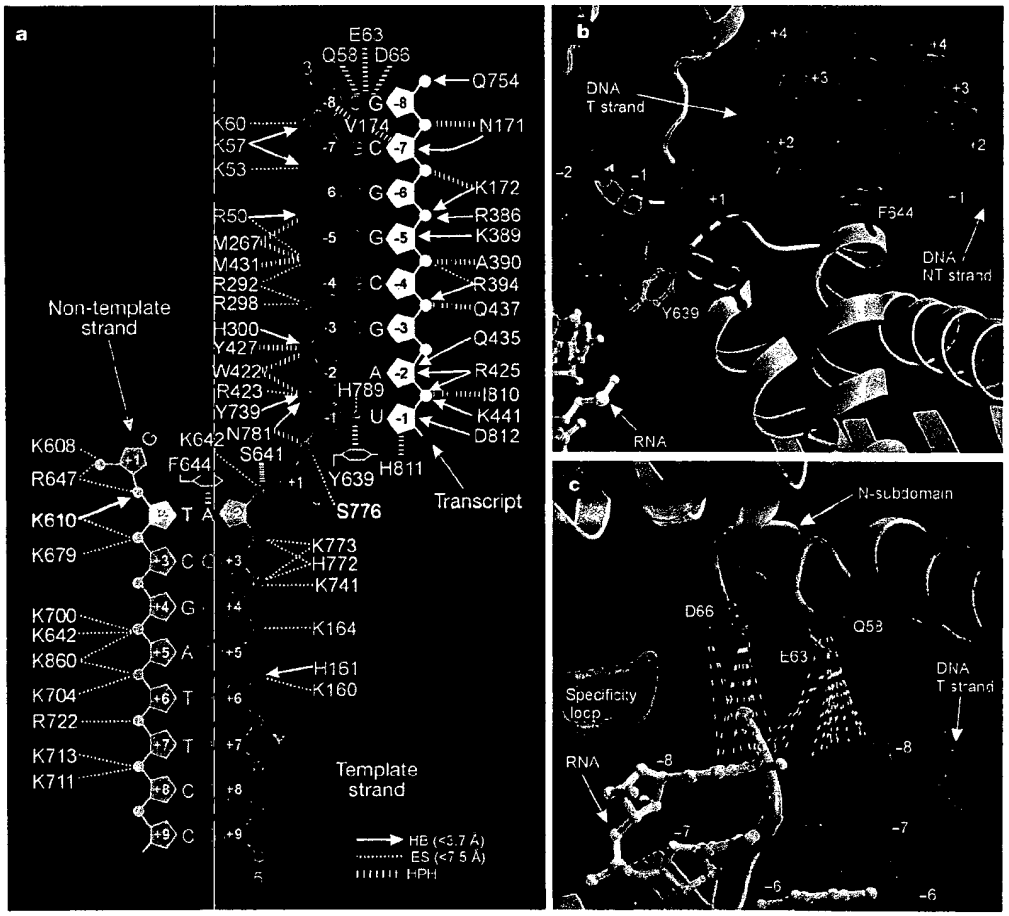


Figure 4 Protein–nucleic acid interactions. **a**, Schematic drawing of the protein–nucleic acid interface. HB, hydrogen bonds; ES, electrostatic interactions; HPH, hydrophobic interactions and closely approaching residues with the nucleotide indicated (including phosphate and sugar). **b**, Junction between downstream DNA and the RNA–DNA hybrid. **c**, Upstream end of the RNA–DNA hybrid. Colours are the same as in Fig. 2. The side chains of Tyr 639 and Phe 644 (**b**), and Gln 58, Glu 63 and Asp 66 (**c**) are shown in orange. Proximity of the N-subdomain residues to the upstream hybrid base pair (**c**) are indicated by dashed lines. T, template; NT, non-template.

of the C-terminal portion of the enzyme (291–303). A final step in the transition from the IC to EC occurs when 10–14 nucleotides of RNA have been synthesized[31,32], and may result from interactions of the displaced RNA with the surface of the pore, thereby stabilizing the transcription bubble.

Another pore provides access to the active site, and is probably the pathway for entering substrate (Fig. 1c). Although this opening is also present in the IC, it becomes more accessible in the EC as a result of a 12° bend of the thumb domain α-helix (372–409) (Fig. 1d). Substitutions of residues that line the surface of the pore result in marked effects on catalytic activity[33–36], which would be consistent with a role for this region in providing a path for entering substrate, or, owing to its proximity to the active site, on catalysis.

The presence of these pores in the EC provides a clear parallel to the multisubunit RNAPs. However, whereas these exit and entry pathways are short in the T7 EC, they are considerably longer (about 30 Å) in the multisubunit RNAPs[15,16]. This difference may account, in part, for the high rate of polymerization exhibited by T7 RNAP (200 nucleotides per s for T7 compared with 30 nucleotides per s for *Escherichia coli* RNAP)[1].

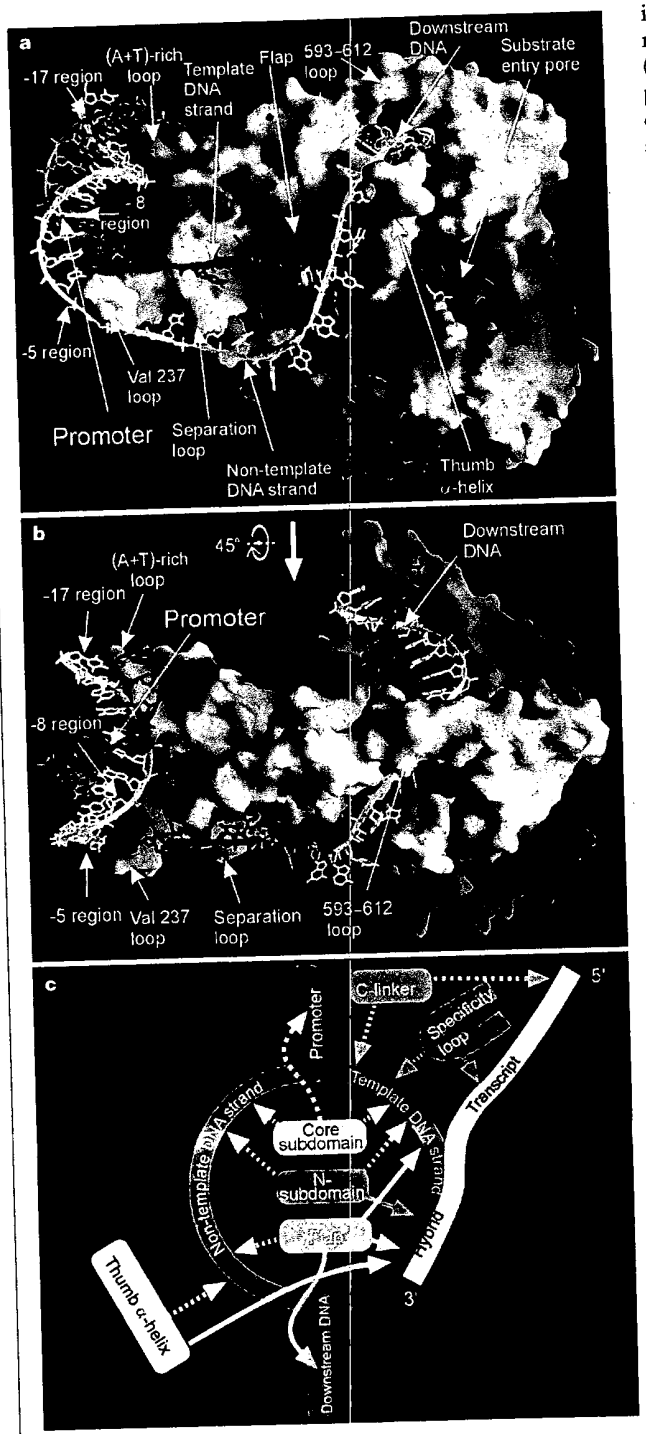


Figure 5 Model of the late IC. Two different views are shown in **a** and **b**. The protein is represented by the same electrostatic potential surface as in Fig. 1c–e. The nucleic acids are shown as a combination of ribbon phosphate backbone and ball-and-stick models with the RNA, and DNA template and non-template strands coloured in light yellow, red and light blue, respectively. **c**, Interaction of the protein domains that have principal roles in the structural organization of the late IC (dashed arrows) or the EC (solid arrows), with the components of the transcription bubble shown schematically. The colour scheme is the same as in Fig. 2.

The transition from an IC to an EC

Biochemical experiments have shown that during the late stages of initiation, just before promoter release, the length of the RNA–DNA hybrid is about 8 bp, as is observed in the EC[7]. It therefore seems likely that the conformation of the RNAP in the late IC resembles that of the EC. A number of questions arise concerning the transition from an early IC to a late IC. Does the transition occur in one step, or is it a multistep process? Furthermore, how are the upstream promoter contacts maintained while the hybrid grows to 8 bp?

Modelling of the RNA–DNA hybrid as it is extended from 3 bp in the early IC indicates that the first clash of the hybrid would occur with elements of the core domain at 4 bp, but that rotation of the core domain, together with specificity loop and promoter, could accommodate extension of the hybrid to 5 bp without substantial refolding (consistent with biochemical data[2,37]). After 5 bp, however, the structure of the EC shows that the hybrid is greatly stabilized by interactions with the refolded elements of the N-terminal domain (Fig. 4a), suggesting that reorganization starts at this point. This is consistent with the observation that mutations affecting the elements in the core domain that interact with the hybrid at this length inhibit extension beyond 5 nucleotides[38,39]. The (A + T)-rich recognition and the Val 237 intercalation loops are part of the unaltered core subdomain, and the promoter contacts mediated by these elements may be retained during the reorientation of this subdomain (assuming that the upstream DNA moves along with the core) and might persist up to 8–9 bp. On the other hand, modelling suggests that this reorganization cannot occur without loss of contacts between the specificity loop and the promoter (Fig. 2a). This is consistent with the results of ultraviolet laser crosslinking experiments, which show that promoter contacts at −7 and −8 are lost before the contact at −17 (ref. 5).

The pattern described above suggests that a simple superposition of the IC and EC core subdomains could provide a plausible model for continued binding of the upstream promoter region in the late IC (Fig. 5). The model shows that a transcription bubble of 9–10 nucleotides (which corresponds to a hybrid length of 5–6 bp) would be too short to connect the promoter in the late IC with the hybrid and downstream DNA (not shown). Thus, the rearrangement that is proposed to commence at 6 bp cannot occur as a single step, and is probably a multistep process in which gradual reorganization of the N-terminal domain is accompanied by incremental movement of its core. The 13–14 nucleotide length of the bubble in the late IC (which corresponds to a hybrid of 8–9 bp, and is in agreement with recent fluorescence experiments[7]) suggests a probable pathway for the template and non-template strands that connect the upstream region of the promoter with the hybrid and downstream DNA (Fig. 5). A 25 Å-long positively charged groove, formed by the thumb domain α-helix and the flap subdomain, is likely to form a

binding site for approximately 6–8 nucleotides of the downstream non-template DNA strand[29,30] (Fig. 5). At the upstream edge of the transcription bubble there are two hydrophobic cavities that can accommodate 4 bases of template and 4–5 bases of non-template strand DNA (Fig. 5). The central wall separating these two cavities (the separation loop) is formed by a portion of the core domain (128–133), whereas the remaining portions are formed by newly formed structural elements. If the movement of the core domain is fixed at this point, further extension of the RNA–DNA hybrid would result in promoter release as a consequence of the strain associated with extension, and/or as a result of compression and extrusion of the template and non-template strands from the hydrophobic channels. Biochemical data suggest that promoter release may precede the final collapse of the transcription bubble and displacement of the 5′ end of the RNA from the hybrid[7]. Interaction of the separation loop with the template and non-template strands just downstream of the promoter may prevent complete collapse of the bubble until later in the initiation process. In the model, the non-template base at −11, which would later re-anneal to the template strand to form the upstream edge of the bubble after promoter release, is on the protein surface in close proximity to the upstream edge of the RNA–DNA hybrid. This suggests that the final transition to the EC may occur without further major alterations of protein structure.

Structural organization of the EC

The EC structure, in combination with modelled promoter DNA and a plausible trace of the template and non-template strands, comprises a transcription intermediate that corresponds to the late IC. Although the modelled nucleic acids are not suitable for detailed analysis of the interactions, they allow a discussion of the role and significance of new structural elements that are acquired by the RNAP in the course of refolding.

The EC structure shows that the rearranged segments of the N-terminal domain (N- and flap-subdomains, and C-linker) are principal elements of the complex. These motifs, which do not have an essential functional role in the early IC, become multi-functional in the late IC and in the EC, and are extensively involved in interactions with various components of the transcription bubble (Fig. 5c).

The N terminus of the N-subdomain probably provides a pathway for the template and non-template DNA strands in both the late IC and the EC. The long α-helix of the N-subdomain interacts with the RNA–DNA hybrid in the EC and may also contribute to the formation of the upstream template and non-template DNA-binding sites in the late IC (Fig. 5c). Importantly, the C-terminal linker region of the N-subdomain is in close proximity to the upstream boundary of the hybrid in the EC, and extension of the hybrid by 1 bp would clash with the hydrophilic residues clustered at the tip of this loop (Fig. 4a, c). The loop may therefore act as a 'hybrid zipper' to facilitate strand separation of the hybrid. In bacterial RNAPs, the σ-subunit 'destabilizer' loop, which is also likely to interact with the upstream edge of the hybrid, contains two highly conserved acidic residues at the tip that may have essentially the same role[19].

In the EC structure, the flap subdomain interacts both with the upstream end of the hybrid and with the downstream DNA. In the late IC, it is probably involved in binding of the upstream portion of the template strand. The flap subdomain also forms part of a potential non-template binding site in both the late IC and the EC (Fig. 5c). Finally, in the EC, this subdomain is required to fix the position of the specificity loop (Fig. 2a), which is indispensable for the formation of the RNA exit channel. Thus, the flap subdomain seems to be of central importance to the integrity and function of the late IC and the EC. Consistent with this, mutations within this motif result in a failure to resolve the transcription bubble[31,40–42].

Although very short (8 residues), the C-linker also probably interacts with two distinct elements of the transcription bubble: with the upstream DNA template strand in the late IC and with the displaced RNA product in the RNA exit pore (Fig. 5c).

Other than the N-terminal domain, the principal element whose functional purpose is switched during the transition to an EC is the specificity loop. While losing its role in promoter recognition it forms a portion of the RNA exit channel and is probably involved in interactions with the displaced RNA product. The flexible out-looped segment of the loop, which is rich in hydrophilic residues, may intrude into the RNA exit pore. We speculate that this segment may be important for responding to pausing and termination signals. The tip of the specificity loop may also interact with the DNA template strand in the late IC (Fig. 5c).

Abortive cycling

Four structures of T7 RNA have been determined previously: free enzyme; RNAP bound to the transcription inhibitor T7 lysozyme; a binary RNAP–promoter complex; and an early IC[11–14]. All of these structures display very similar conformation of the enzyme (Fig. 2b), which implies that this conformation probably represents the most stable organization of the protein. In contrast, the conformation of RNAP in the EC (Fig. 2a) has never been observed before. The extensive protein–nucleic acid interface (about 2,500 Å^2) in the EC, which is lacking in the other structures, suggests that the protein conformation of the EC is greatly stabilized by these interactions. During the transition to an EC the exposed hydrophobic surface of the N-terminal domain increases by roughly 1,400 Å^2, which, in the absence of the nucleic acid–protein interface, may favour a return to the more stable IC conformation. Given the probable multistep nature of the transition from the early IC to the late IC, it seems probable that the intermediate enzyme conformations would be less stable than the one observed in the EC, because they lack the full extent of the final nucleic acid interface. In addition, modelling shows that the principal N-subdomain and C-linker α-helices can only be gradually folded during the course of the transition, and that the unfolded IC loop, comprising the flap domain in the EC, would lose its stabilizing interactions with the protein during the initial stage of the transformation. We speculate that competition between the nascent RNA–DNA hybrid for space in the active site (which induces the reorganization of the RNAP) against the tendency of the partially folded protein structures to fold back to the initial conformation is a major factor contributing to abortive cycling. A similar mechanism has been proposed for bacterial multisubunit enzymes, in which the growing RNA–DNA hybrid competes with initiation factor-σ resulting either in abortive RNA synthesis or displacement of σ to form a stable EC[18,19].

Discussion

On the basis of footprinting experiments, two principal models were proposed for the progress of transcription initiation: 'polymerase inchworming', in which the RNAP structure extends to include the expanding transcription bubble, and 'DNA scrunching', in which the RNAP structure remains unaltered while the bubble is accumulated in the active site. The interpretation of previous structural data favoured the scrunching model, in which the RNA–DNA hybrid length was limited to 3 bp and the strain associated with packing of the transcription bubble in the unaltered enzyme led to abortive initiation[14].

Although not excluding a role for scrunching during the early stages of transcription (when the hybrid is ≤4 bp), the T7 RNAP EC structure reported here provides strong evidence in favour of polymerase inchworming during the later stages of initiation, as the structure of the RNAP is expected to change progressively to accommodate the RNA–DNA hybrid as it grows to a final length of about 8–9 bp.

During the maturation of the EC, the elaboration of the components of the transcription bubble both induce and stabilize the reorganization of the protein structure. We suggest that a similar

role for nucleic acids may also be expected for other multifunctional nucleic-acid-binding proteins, including the multisubunit RNAPs during their transition from an IC to an EC.

The overall organization of the T7 RNAP EC bears a marked resemblance to that of the multisubunit RNAPs. Moreover, the structural mechanism by which T7 RNAP achieves this final state is similar to the steps that are observed in bacterial RNAPs. In both cases the active site that accommodates the components of the mature transcription bubble is greatly reduced at the beginning stages of transcription, and the RNA exit channel is either not formed or is blocked. These elements probably form gradually in a process that involves competition between the growing RNA–DNA hybrid and protein structural elements. The reorganization observed in T7 RNAP therefore resembles the displacement and release of the bacterial σ-factor, and we suggest that a similar situation may exist in the multisubunit eukaryotic RNAPs. The induced reorganization of the enzyme as it matures and makes the transition to an EC is therefore probably a consistent theme among DNA-dependent RNAPs. □

Methods

Crystallization and data collection

Stable, transcriptionally competent complexes were formed by mixing RNAP at a 1:1 molar ratio with a nucleic acid scaffold consisting of an 18-nucleotide template DNA strand annealed to 10 nucleotides of downstream DNA and a 12-nucleotide RNA primer, of which 4 nucleotides at the 5′ end were unpaired with the template. Such complexes exhibit all of the properties of a promoter-initiated EC[24].

Crystallization was carried out by the sitting-drop vapour diffusion technique at 20 °C (ref. 25). The drop, containing 2 μl of 10 mg ml^{-1} complex solution, was mixed with 2 μl of well solution containing 10% PEG 8000, 10% glycerol, 5 mM β-mercaptoethanol, 100 mM Tris buffer, pH 8.1. Thin, plate-like crystals (0.5 × 0.5 × 0.05 mm) were grown within 3–5 days and diffracted to 2.6 Å resolution using synchrotron X-ray radiation. The T7 RNAP EC crystals belong to the *P*1 space group with unit cell dimensions $a = 79.91$, $b = 84.97$, $c = 202$ Å, $\alpha = 90.36$, $\beta = 92.97$, $\gamma = 109.94°$. A complete diffraction data set was collected at 100 K on SPring-8 beam line BL45XU, and processed with the HKL2000 program package[43] (Supplementary Table 1).

Structure determination and refinement

The structure was determined by molecular replacement using the coordinates of the T7 RNAP IC (Protein Data Bank accession number 1QLN[14]) as a search model[44]. Clear molecular replacement solutions for four crystallographically independent molecules were obtained only after removal of the N-terminal domain (266 residues) and all protruding segments of the C-terminal domain from the initial model, which indicated large structural alterations of the omitted portions of the molecule[23]. The phases obtained from the initial structure were improved by density modification and non-crystallographic symmetry (NCS) averaging using the DM program[44], resulting in an interpretable electron density for the omitted N-terminal domain and the RNA–DNA hybrid. The atomic model was built on the basis of this density using the O program[45] and then refined by the CNS program[46]. At this stage, however, the refinement was unstable, resulting in large portions of density that disappeared from the ($2F_o - F_c$) electron density maps after additional refinement cycles. Careful inspection of diffraction data revealed barely separated twin spots at medium resolution, which become more profound at high resolution. This was characteristic of all crystals examined. To avoid this problem, the size of integration spots was increased at the stage of data processing to include both portions of the spots twinned at high resolution, and the diffraction intensities were calculated using summation rather than profile fitting to allow for better evaluation of the intensities for the two peaks in the integrated area. The four-fold NCS averaging appeared to be a powerful tool for the evaluation of the quality of the diffraction data (see Supplementary Information section 'NCS averaging'). The reprocessed data at 2.9 Å resolution resulted in clear electron density, which allowed for unambiguous modelling of the missing regions of the molecules with almost all protein side chains. The four independent molecules in the asymmetric unit are almost identical, with r.m.s.d. between the molecules ranging from 0.77 to 1.12 Å for the protein main chain and nucleic acid atoms. The downstream DNA is flexible in two molecules, and two regions of the protein were not resolved (757–765 and 352–371). Only weak electron density was visible for the four unpaired nucleotides (AACU) in the 5′ region of the RNA and for the last base pair of the downstream DNA (G•C, position +10). Figures 1a, b, 2a, b, 3 and 4b, c, were prepared using the programs Molscript[47], Bobscript[48] and Raster3D[49]. Figures 1c–e and 5 were prepared using the GRASP program[50].

Received 29 July; accepted 19 September 2002; doi:10.1038/nature01129.
Published online 9 October 2002.

1. McAllister, W. T. Transcription by T7 RNA polymerase. *Nucleic Acids Mol. Biol.* **11**, 15–25 (1997).
2. Brieba, L. G. & Sousa, R. T7 promoter release mediated by DNA scrunching. *EMBO J.* **20**, 6826–6835 (2001).
3. Gunderson, S. I., Chapman, K. A. & Burgess, R. R. Interactions of T7 RNA polymerase with T7 late promoters measured by footprinting with methidiumpropyl-EDTA-iron(II). *Biochemistry* **26**, 1539–1546 (1987).
4. Ikeda, R. A. & Richardson, C. C. Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription. *Proc. Natl Acad. Sci. USA* **83**, 3614–3618 (1986).
5. Place, C., Oddos, J., Buc, H., McAllister, W. T. & Buckle, M. Studies of contacts between T7 RNA polymerase and its promoter reveal features in common with multisubunit RNA polymerases. *Biochemistry* **38**, 4948–4957 (2000).
6. Martin, C. T., Muller, D. K. & Coleman, J. E. Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* **27**, 3966–3974 (1988).
7. Liu, C. & Martin, C. T. Promoter clearance by T7 RNA polymerase. *J. Biol. Chem.* **277**, 2725–2731 (2002).
8. Liu, C. & Martin, C. T. Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. *J. Mol. Biol.* **308**, 465–475 (2001).
9. Temiakov, D. *et al.* The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc. Natl Acad. Sci. USA* **97**, 14109–14114 (2000).
10. Huang, J. & Sousa, R. T7 RNA polymerase elongation complex structure and movement. *J. Mol. Biol.* **303**, 347–358 (2000).
11. Sousa, R., Chung, Y. J., Rose, J. P. & Wang, B. C. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* **364**, 593–599 (1993).
12. Jeruzalmi, D. & Steitz, T. A. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J.* **17**, 4101–4113 (1998).
13. Cheetham, G., Jeruzalmi, D. & Steitz, T. A. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399**, 80–83 (1999).
14. Cheetham, G. & Steitz, T. A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309 (1999).
15. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
16. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. *Science* **292**, 1863–1876 (2001).
17. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Ångstrom resolution. *Science* **292**, 1876–1882 (2001).
18. Murakami, K. S., Musada, S. & Darst, S. A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**, 1280–1284 (2002).
19. Vassylyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å reslolution. *Nature* **417**, 712–719 (2002).
20. Cheetham, G. & Steitz, T. A. Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* **10**, 117–123 (2000).
21. Severinov, K. T7 RNA polymerase transcription complex: what you see is not what you get. *Proc. Natl Acad. Sci. USA* **98**, 5–7 (2000).
22. Steitz, T. A. DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3**, 31–38 (1993).
23. McAllister, W. T. & Raskin, C. A. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol. Microbiol.* **10**, 1–6 (1993).
24. Temiakov, D., Anikin, M. & McAllister, W. T. Characterization of T7 RNA polymerase transcription complexes assembled on nucleic acid scaffolds. *J. Biol. Chem.* (in the press).
25. Temiakov, D. *et al.* Crystallization and preliminary crystallographic analysis of T7 RNA polymerase elongation complex assembled on an RNA:DNA scaffold. *Acta Crystallogr.* (submitted).
26. Kim, Y. *et al.* Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* **376**, 612–616 (1995).
27. Raskin, C. A., Diaz, G. A. & McAllister, W. T. T7 RNA polymerase mutants with altered promoter specificities. *Proc. Natl Acad. Sci. USA* **90**, 3147–3151 (1993).
28. Rong, M., He, B., McAllister, W. T. & Durbin, R. K. Promoter specificity determinants of T7 RNA polymerase. *Proc. Natl Acad. Sci. USA* **95**, 515–519 (1998).
29. Ma, K., Temiakov, D., Jiang, M., Anikin, M. & McAllister, W. T. Major conformational changes occur during the transition from an initiation complex to an elongation complex by T7 RNA polymerase. *J. Biol. Chem* (in the press).
30. Mukherjee, S., Brieba, L. G. & Sousa, R. Structural transitions mediating transcription initiation by T7 RNA polymerase. *Cell* **110**, 1–20 (2002).
31. Mentesana, P. E., Chin-Bow, S. T., Sousa, R. & McAllister, W. T. Characterization of halted T7 RNA polymerase elongation complexes reveals multiple factors that contribute to stability. *J. Mol. Biol.* **302**, 1049–1062 (2000).
32. Jiang, M., Rong, M., Martin, C. T. & McAllister, W. T. Interrupting the template strand of the T7 promoter facilitates translocation of the DNA during initiation, reducing transcript slippage and the release of abortive products. *J. Mol. Biol.* **310**, 509–522 (2000).
33. Bonner, G., Lafer, E. M. & Sousa, R. Characterization of a set of T7 RNA polymerase active site mutants. *J. Biol. Chem.* **269**, 25120–25128 (1994).
34. Huang, J., Brieba, L. G. & Sousa, R. Misincorporation by wild type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **39**, 11571–11580 (2000).
35. Osumi-Davis, P. A., de Aguilera, M. C., Woody, R. W. & Woody, A. Y. Asp537, Asp812 are essential and Lys631, His811 are catalytically significant in bacteriophage T7 RNA polymerase activity. *J. Mol. Biol.* **226**, 37–45 (1992).
36. Woody, A. Y., Osumi-Davis, P. A., Hiremath, M. M. & Woody, R. W. Pre-steady state and steady-state kinetic studies on transcription initiation catalyed by T7 RNA polymerase and its active site mutants K631R and Y639F. *Biochemistry* **37**, 15958–15964 (1999).
37. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **39**, 10419–10430 (2000).
38. He, B., Rong, M., Durbin, R. K. & McAllister, W. T. A mutant T7 RNA polymerase that is defective in RNA binding and blocked in the early stages of transcription. *J. Mol. Biol.* **265**, 275–288 (1997).
39. Brieba, L. G., Gopal, V. & Sousa, R. Scanning mutagenesis reveals roles for helix N of the bacteriophage T7 RNA polymerase thumb subdomain in transcription complex stability, pausing, and termination. *J. Biol. Chem.* **276**, 10306–10313 (2000).

40. Lyakhov, D. L. *et al.* Mutant T7 RNA polymerases with altered termination properties. *J. Mol. Biol.* **269**, 28–40 (1997).
41. Macdonald, L. E., Durbin, R. K., Dunn, J. J. & McAllister, W. T. Characterization of two types of termination signal for bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **238**, 145–158 (1994).
42. Gopal, V., Brieba, L. G., Guajardo, R., McAllister, W. T. & Sousa, R. Characterization of structural features important for T7 RNAP elongation complex stability reveals competing complex conformations and a role for the non-template strand in RNA displacement. *J. Mol. Biol.* **290**, 411–431 (1999).
43. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
44. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
45. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
46. Brunger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
47. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
48. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
49. Merrit, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
50. Nichols, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Supplementary information accompanies the paper on *Nature*'s website (**http://www.nature.com/nature**).

Acknowledgements We thank Y. Kawano for assistance during the data collection at the SPring-8 synchrotron beam line, BL45. We are grateful to A. Murzin for discussions and advice concerning the analysis of the structure. This work was supported in part by grants from the NIH (USA) (W.T.M.) and the Organized Research Combination System of Science and Technology Agency (Japan) (S.Y.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.G.V. (e-mail: dmitry@yumiyoshi.harima.riken.go.jp) or S.Y. (e-mail: yokoyama@biochem.s.u-tokyo.ac.jp). The atomic coordinates have been deposited in the Protein Data Bank under accession number 1H38.

Sciencexpress Research Article

Structural Basis for the Transition from Initiation to Elongation Transcription in T7 RNA Polymerase

Y. Whitney Yin[1] and Thomas A. Steitz[1,2,3]*

[1]Department of Molecular Biophysics and Biochemistry, [2]Department of Chemistry, [3]Howard Hughes Medical Institute, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520–8114, USA.

*To whom correspondence should be addressed. E-mail: eatherton@csb.yale.edu

To make mRNA transcripts, bacteriophage T7 RNA polymerase (T7 RNAP) undergoes a transition from an initiation phase, which only makes short RNA fragments, to a stable elongation phase. We have determined at 2.1 Å resolution the crystal structure of a T7 RNAP elongation complex with 30 base pairs of duplex DNA containing a "transcription bubble" interacting with a 17 nucleotide RNA transcript. The transition from an initiation to an elongation complex is accompanied by a major refolding of the N-terminal 300 residues. This results in loss of the promoter binding site, facilitating promoter clearance, and creates a tunnel that surrounds the RNA transcript after it peels off a 7 base pair heteroduplex. Formation of the exit tunnel explains the enhanced processivity of the elongation complex. Downstream duplex DNA binds to the fingers domain and its orientation relative to upstream DNA in the initiation complex implies an unwinding that could facilitate formation of the open promoter complex.

Despite structural differences, the 99 kDa single subunit RNA polymerase from the bacteriophage T7 (T7RNAP) and the multi-subunit cellular RNA polymerases share numerous functional characteristics. Both families of RNA polymerases have initiation and elongation phases of transcription (*1*). During the initiation phase, RNAP binds to a specific DNA promoter, opens the duplex at the transcription start site and initiates RNA synthesis de novo. Transcription during this phase is unstable and characterized by repeated abortive initiation events that produce short RNA fragments (2-6 nucleotides) (*2*, *3*). After synthesis of 10 to 12 nucleotide long RNA, the polymerase enters the elongation phase and completes transcription of the mRNA processively without dissociating until termination. There are significant biochemical differences between the initiation and elongation states. Footprinting assays show differences in DNA protection (*4–6*). The T7RNAP-DNA complex is significantly more stable in the elongation phase (*3*, *7*), and T7 lysozyme, a natural inhibitor of T7RNAP, inhibits the transition from the initiation to elongation state, but has little effect on the activity of the transcribing elongation complex (*8*). Here, we describe the structural basis for promoter opening , the transition from the abortive initiation to processive elongation phases, promoter clearance, the regulation by T7 lysozyme and the unwinding of downstream DNA.

The structure of T7RNAP was largely unchanged when complexed either with the transcription inhibitor T7 lysozyme (*9*), a 17 base pair open promoter DNA (*10*) or a 17 base pair promoter containing a 5' template extension of 5 nucleotides and a 3 nucleotide RNA transcript (*11*); the C-terminal two thirds of T7 RNAP is homologous to the polymerase domain of the pol I family DNA polymerase (*12–15*) while a novel N-terminal domain (residues 1-325) is unique to the RNA polymerase. The structures show that one antiparallel β loop, named the specificity loop (residues 740-770), makes sequence-specific contacts with the promoter while another, the intercalating hairpin (residues 230-240), opens the upstream end of the transcription bubble (*10*, *11*). The structures of the two promoter containing complexes also provide support for the 'scrunching' model of transcription initiation in which RNA synthesis leads to an accumulation of the DNA template within the active site before the promoter is released (*11*).

Nevertheless, the structure of the transcribing initiation complex could not explain many aspects of the elongation phase. Extensive proteolysis that results in loss of the N-terminal 180 residue fragment abolishes the elongation phase; although the polymerase can initiate transcription from a promoter, it makes only abortive transcripts (*3*, *16*). A proteolytic cleavage of T7RNAP after residue 173 or 180 results in somewhat decreased efficiency of elongation and decreased single-strand RNA binding, which suggests that the integrity of the region between residues 170-180 plays a role in elongation (*3*). However, since the proteolytically nicked region is located at least 40 Å away from the 5′ end of the RNA transcript in the initiation complex structure, it was not clear how this remote site could affect elongation transcription. Similarly, a single mutation, E148A, which abolished synthesis of any transcript longer than 5 nucleotides (*17*), is located at least 35 Å away from the 5′ end of the mRNA in the initiation complex structures.

The most puzzling paradox, however, arose from the apparent incompatibility of the biochemical and structural evidence for the maximum length of the DNA-RNA heteroduplex during transcription. The structure of the initiation complex contained only 3 base pairs of DNA-RNA heteroduplex, and indeed any extension of the heteroduplex beyond three base pairs was deemed to be sterically excluded by the protein structure (*11*). In contrast, numerous biochemical studies, including a recent crosslinking of the RNA and DNA strands, led to the conclusion that the length

of the DNA-RNA heteroduplex during elongation is approximately eight base pairs (*6, 18, 19*). Various attempts (*19*) to accommodate these data using the existing T7RNAP structures seemed implausible. We show here that these apparent conflicts arose because an important piece of the puzzle in understanding the transition from transcription initiation to elongation by T7RNAP was missing.

The crystal structure of a T7RNAP complex trapped in a functional elongation mode with a transcription bubble of DNA and heteroduplex RNA seven nucleotides long shows that the N-terminal domain changes conformation dramatically compared to the initiation complex. As a consequence of this change, the promoter binding site is destroyed and a channel that accommodates the heteroduplex in the active site and an exit tunnel through which RNA can pass are created. These structural features account for promoter clearance and processivity in the elongation phase.

Structure of an elongation complex. The T7RNAP was co-crystallized with a DNA containing a transcription bubble and mRNA, a complex that mimics the elongation phase, and its structure was determined at 2.1 Å resolution. The 30 base pair duplex DNA contains a central region of 11 non-complementary bases and a 17 nucleotide RNA that is complementary to the template for 10 nucleotides at its 3' end (Fig. 1A). The RNA of this substrate can be extended by the polymerase in a template-directed and processive manner in the absence of promoter (*20*), and it possesses other features of a normal promoter-initiated elongation complex, as seen in the earlier work of von Hippel and Daube(*21*). The T7RNAP elongation complex was assembled by mixing the polymerase with the substrate after the three oligonucleotide strands of template DNA, non-template DNA and RNA had been annealed (*22*). The structure was determined by single wavelength anomalous diffraction using selenomethionine-substituted T7RNAP and by molecular replacement using the polymerase domain of the T7RNAP initiation complex as a search model. The phases derived from these two sources were weighted and combined. Density modification was applied to the initial electron density map calculated with combined phases to further reduce the phase errors and improve the map (Fig. 1B) (*23*). The final refined model has an R_{factor} of 24.1% (R_{free} = 27.5%). The data collection and refinement statistics are provided in Table 1.

The T7RNAP protein is seen bound to duplex DNA and its RNA transcript annealed to a central, opened section of DNA in the active site (Fig. 1B). The active site is located in an enlarged channel bounded by the polymerase's fingers, thumb and palm domains of the C-terminus. Although the annealed construct contained ten nucleotides of complementary heteroduplex RNA-DNA by design, the elongation complex active site contains only seven base pairs of heteroduplex DNA-RNA. After seven base pairs, the 5' end of the mRNA is separated from the template by an a helix of the thumb domain and enters a positively charged tunnel in the protein while the template strand remains bound to the thumb domain. The single-stranded, non-template DNA in the bubble is separate from the template DNA-RNA heteroduplex and makes sequence-independent contacts with the protein. The template and non-template strands merge to form duplex DNA at both the upstream and downstream ends of the bubble (Fig. 1B). Only three nucleotides of DNA upstream of the DNA-RNA heteroduplex are visible in the electron density map. At the downstream end of the bubble the non-template DNA strand base pairs with the template strand at a position that is one nucleotide beyond the incoming nucleotide ($n + 1$), and all 10 downstream base pairs are clearly visible in a positively charged cleft formed by the fingers domain. Overall, the 3' terminal 10 nucleotides of RNA as well as 22 nucleotides each of template and non-template DNA are visible in the electron density maps, while eight base pairs of upstream DNA duplex and the 5' terminal seven nucleotides of single stranded RNA are not (Fig. 1A). The seven base pair RNA-DNA heteroduplex seen in this structure is in good agreement with the ~8 base pair length derived by biochemical studies (*6, 18, 19*). The total of 22 nucleotides of DNA seen in the structure also agrees with the 22-24 nucleotides length of DNA protected by the polymerase in foot-printing experiments of an elongation complex (*4, 6, 16*).

This elongation complex is analogous to the binary complexes of primer-template DNA with DNA polymerases; its primer terminus is located at the post-translocation position ready to accept an incoming nucleotide. In this complex, as with the corresponding DNA polymerase complexes (*24*), the base that is to form the template for the incoming nucleotide (n) lies in a pocket in the fingers domain, rather than in an orientation that would allow it to pair with the incoming nucleoside triphosphate. The position of the primer terminus relative to the palm domain also is identical to its position in the transcription initiation complex lacking the NTP (*10*). We have additionally determined the structure of an elongation complex after insertion of the nucleotide at the n position yielding an 8 base pair heteroduplex and insights into the mechanism of translocation (Yin and Steitz, in preparation).

T7 RNAP structural transition from initiation to elongation. Comparison of the polymerase structure in the initiation complex with its structure in the elongation complex shows that portions of the enzyme, most notably the N-terminal domain, have undergone significant conformational changes that alter its shape and tertiary structure (Fig. 2). The structural changes seen in the N-terminal domain involve three different regions, each undergoing a different kind of conformational alteration: 1) Rigid body motion. Six α-helices, D,E,F,G and I,J, rotate by 140° and translate by 30 Å as a rigid body from their location in the initiation complex (Fig. 2F). The ends of helices F and G pack against the third base pair of heteroduplex in the initiation complex and must move to accommodate a longer heteroduplex. The six helices are repositioned into the region that is occupied by the promoter DNA in the initiation complex, thereby abolishing the interaction between T7RNAP and the promoter, and thus explaining promoter clearance. The intercalating hairpin, which opens the promoter in the initiation complex, moves and becomes disordered in the elongation complex, accounting for mutations in this loop that reduce the efficiency of initiation but not elongation (*25*). 2) Extension of an α-helix. In a

conformational change that is strikingly reminiscent of one undergone by the flu virus hemagglutinin protein upon a pH change (*26*), helix C1 becomes significantly elongated from 22 Å to 50 Å by the stacking of helix C2 on top of C1 and the refolding of the disordered loop between C2 and helix D to further elongate the C1-C2 helix (Fig. 2E). The elongated C-helix now protrudes into the region formerly occupied by the six-helix assembly in the initiation complex, implying that the two conformational changes are likely to be coordinated. 3) Formation of anti-parallel α-helices. Perhaps the most dramatic and unprecedented conformational change involves residues 160 to 190 which not only extensively refold, but move about 70 Å from one side of the polymerase to the other (Fig. 2G). This region refolds from a short helix and an extended loop into a pair of anti-parallel helices (H1,H2/3) (Fig. 2A,B,F). The newly formed compact structure, named sub-domain H, forms part of the RNA-transcript exit tunnel and contacts the 5′ end of the RNA transcript on one surface and the non-template DNA on the opposite surface.

In contrast with the N-terminal domain, the C-terminal domain undergoes fewer structural changes (Fig. 3B). The thumb domain rotates by about 15° from its orientation in the initiation complex, and together with the sub-domain H, creates a binding cleft for the non-template strand DNA (Fig. 3C,D). In the initiation complex, the specificity loop crosses the active site to make sequence specific interactions in the major groove of the promoter DNA (*10, 11*), and it lies in a position that blocks the path through which RNA exits in the elongation complex. In the elongation complex the specificity loop moves sideways to open the exit tunnel and become a part of it. The tip of the specificity loop, which contacts promoter DNA in the initiation complex, now contacts the 5′ end of the RNA message in its passage through the tunnel (Fig. 3D,4), consistent with the observation that the RNA transcript can be photocross-linked to the specificity loop (*19*). This conformational change of the specificity loop may also be associated with promoter release.

The mutation E148A, which lies remote from substrates in the initiation complex, abolishes synthesis of transcripts longer than 5 nucleotides (*17*). This may be due to the inability of the mutant 148A to make interactions necessary to the structural transition of the specificity loop in forming an elongation complex. E148 stacks directly against M750 and interacts indirectly with N748 at the tip of the specificity loop through R155 to bend the specificity loop toward the 5′ end of the RNA. The mutant 148A cannot make the interactions to secure the bending configuration of the specificity loop and therefore would affect the integrity of the exit tunnel. Consequently, both structural and biochemical studies agree on these dual functions of the specificity loop.

This massive structural reorganization of the N-terminal domain upon formation of the elongation complex creates a tunnel through which the RNA can exit and a binding site for the single stranded non-template DNA of the transcription bubble from *n- 7* to *n*. The newly formed exit tunnel, whose interior is positively charged, measures approximately 8 Å in diameter and 20 Å in length and is formed by the thumb domain, the specificity loop and sub-domain H (Fig. 3C,D and 4). After seven base pairs of heteroduplex, three nucleotides of RNA are separated from the DNA by the rim of the exit tunnel (residues 170-180) and the thumb domain (Fig. 3C). The single stranded 5′end of the RNA transcript is seen entering the exit tunnel (Fig. 4B,D). Model building suggests that the tunnel may accommodate five extended nucleotides, implying that an RNA transcript longer than 12 nucleotides would emerge from the side of the tunnel opposite the active site. The exit tunnel contacts the RNA strictly through interactions with the phosphates of the sugar-phosphate backbone. Residues Arg756 and Gln754 from the specificity loop, as well as Asn171 and Lys172 on the sub-domain H are all within hydrogen bonding distance of phosphates at the 5' end of the single stranded mRNA (Fig. 3D).

The processivity of the elongation complex, in contrast to the initiation complex, could be explained by the appearance of the mRNA exit tunnel, which topologically surrounds transcripts longer than 10-12 nucleotides. Its effect on processivity is entirely analogous to that of the sliding clamp on DNA replication (*27*). Further, the stability of the elongation complex, compared to the abortive phase complex, is enhanced by the extensive interaction between the seven base pairs of heteroduplex and its binding site. That proteolytic cleavage of residues between 170 to 180 results in reduced elongation efficiency (*3, 16*), may be a consequence of a reduced integrity of domain H and stability of the tunnel in the elongation phase. Complete proteolytic digestion of the N-terminal 180 residues results in a T7RNAP capable of abortive synthesis and incapable of elongating transcripts beyond eight nucleotides (*3, 15*). As this enzyme would be missing helices D-G and subdomain H, it may be incapable of destroying the promoter binding site, which is required for clearance, and unable to form the RNA exit tunnel required to processive synthesis.

During the transition from initiation to elongation, T7RNAP relinquishes its sequence-specific grasp of the promoter and begins translocation along DNA, a process often referred to as promoter clearance, which is achieved by the destruction of the promoter binding site and movement of the six helices by 30 Å into the position formerly occupied by the promoter DNA (Fig. 2). Upstream DNA now binds in a sequence independent manner to a newly created cleft that is formed in part by the thumb domain and helix C2 (Figs. 2B, 4B). These two upstream DNA binding sites are separated by at least 40 Å and cannot be occupied simultaneously by upstream DNA since formation of one binding site dismantles the other. The DNA that is upstream of the transcription bubble and visible in our complex is not base paired due to the non-complementarity of the designed sequence. Presumably, the upstream DNA of complementary sequence would form duplex that would lie in the upstream channel.

What causes the T7RNAP to undergo this significant conformational change and what stabilizes the elongation phase structure? Since the apo enzyme has essentially the same structure as the initiation complex structure (*10, 11*), it seems likely that the formation of the longer RNA:DNA heteroduplex is playing a significant role. We previously noted that it was not possible to elongate the heteroduplex beyond three base pairs in the initiation complex due to a

steric clash with helices F and G (*11*). Here, we observe that the three base pair heteroduplex in the initiation complex and the seven base pair heteroduplex in the elongation complex are bound and oriented identically on the polymerase active site (Fig. 3B). When the C_α backbone atoms of the palm domains of the initiation and elongation complexes are superimposed, the first two base pairs of the initiation and elongation complexes superimpose with an RMSD of 0.34 Å (Fig. 3B). Thus, it seems likely that incorporation of a fourth nucleotide into the transcript would result in a steric clash and a destabilization of the initiation complex structure. Indeed, our attempts to elongate the 3 nucleotide transcript by one nucleotide destroyed transcribing crystals of the initiation complex. However, DNA protection experiments suggest that T7RNAP is still bound to the promoter in the presence of a six nucleotide transcript (*4*), which is inconsistent with a complete conversion of T7 RNAP to the elongation complex structure seen here. Taken together, the structural and footprinting data imply either that an additional T7RNAP conformation exists that allows the formation of a longer heteroduplex than can be accommodated by the initiation complex while still making promoter specific interactions, or, alternatively, that the RNA peels off the template after 3 nucleotides, as suggested earlier (*11*), until it becomes long enough to form the seven base pair heteroduplex seen in the elongation complex. It is not obvious what kind of intermediate conformational change would move helices F and G from blocking an elongating heteroduplex while not destroying the promoter binding site, since helices F and G move in concert with helices D-G in the observed transition (Fig. 2). In any case, the longer heteroduplex should destabilize the initiation complex conformation of T7RNAP and make interactions that stabilize the elongation conformation of the enzyme. Indeed, the total surface area of contact between T7 RNAP and the DNA/RNA substrates is the same in the initiation and elongation structures, but only after the heteroduplex reaches 7 b.p. in length.

One might ask why the abortive synthesis of short oligonucleotides exists and why the enzyme might not be "designed" to carry out the stable RNA synthesis that occurs in the elongation phase right from the start. The initiation of RNA synthesis at a particular site that is required for specific gene expression and regulation as well as the need for de novo, unprimed synthesis necessitates binding of the polymerase at a specific DNA location, the promoter. Furthermore, the binding of T7RNAP to both the promoter and downstream DNA appears to be essential for opening the bubble. Since short transcripts (2-4 nt) cannot form stable heteroduplexes, polymerase leaving the promoter prematurely would lead, presumably, to bubble closure and transcript displacement by the non-template strand. An enzyme locked in the elongation mode conformation seems unlikely to be capable of specific initiation and bubble opening.

T7 RNAP opening of the transcription bubble. T7RNAP appears to facilitate formation of a transcription bubble by untwisting and bending duplex DNA. To derive the degree of promoter untwisting and bending upon binding to T7RNAP, a complete open promoter DNA was constructed by superimposing the palm domains of the open promoter and the elongation complexes and combining the upstream DNA from the former with the downstream DNA of the latter. This complete open promoter contained 13 b.p. of upstream promoter duplex, 10 b.p. of downstream duplex and six nucleotides of template between –4 and +2, for a total of 29 nucleotides of template DNA. After superimposing the 13 b.p. promoter on one end of a 29 b.p. B-form duplex, the other end of the straight DNA has to be bent by 80 degrees and untwisted by 146° to superimpose on the downstream DNA of the complete open promoter. The bubble is 6 nucleotides long and includes nucleotides -4 to +2 (promoter numbering). The template strand of the bubble, which is visible in the complexes, bends sharply (~90°) at position -4 and descends into the active site and likewise bends about 80° after +2 to re-emerge from the deeply buried active site and rejoin the downstream duplex at +3.

We suggest that energy required to melt the 6 base pairs of duplex to form the bubble (about 9-16 kcal/mole) may arise from the reduction of the DNA twist by about 146° and changing the relative orientations of the upstream and downstream DNA axes by 80°. The underwinding of DNA produced by promoter binding could destabilize duplex by up to 24 kcal/mol (*28*) while the bending may destabilize it by as much as 25 kcal/mol (*28*), either one of which is sufficient to melt the duplex. It is presumably the extensive interaction between the enzyme and a bent, unwound, open promoter DNA that produces the free energy required to distort the DNA thereby destabilizing and opening the duplex. The total surface area of the initial open promoter DNA that interacts with T7RNAP is 2700 $Å^2$. Using the conversion factor of 25 cal/$Å^2$ of buried surface area, which is often used to evaluate the energetic contributions of hydrophobic interactions to binding (*29*), we can calculate that as much as 68 Kcal of intrinsic interaction energy may be available for DNA distortion, entropy of immobilization reducing conformational entropy and a 10^{-9} M dissociation constant (*30*). In this regard it may be interesting to note that the several insertions in the fingers domain of T7RNAP as compared with Klenow fragment serve to greatly increase the interaction surface with downstream DNA in the RNA polymerases.

Non-template DNA. The single-stranded non-template DNA is well separated from the heteroduplex in the transcription bubble (Figs. 1B,3C), which is held open by extensive interactions between the polymerase and both the template DNA-RNA heteroduplex and the single stranded non-template strand. The non-template DNA is immobilized at two points along the transcription bubble. At the upstream fork, it interacts with the thumb domain and the outer surface of the RNA exit tunnel that is formed by sub-domain H. Bases at positions *n-2* and *n-3* are flipped out of the helical axis and stack with Arg173 of sub- domain H and Tyr385 of the thumb domain, respectively. At the downstream fork, it interacts with the fingers domain. The non-template DNA interacts with one side of sub-domain H while the RNA message interacts with the other (Fig. 3C).

Crystal structures of the initiation complex and the open promoter complex did not show non-template DNA downstream of *–5* (promoter numbering). In both complexes,

the DNA duplex from *-1* to *-4* was melted with the template strand firmly bound and plunging into the active site, while the non-template strand was disordered(*10, 11*). If the non-template strand is modeled into these initiation complexes at the position it occupies in the elongation complex, there are no plausible interactions apparent, as subdomain H lies on the opposite side of the molecule in the initiation complex and the position of the thumb is also altered. This observation agrees with biochemical studies showing that the presence of non-template DNA in the bubble region stabilizes the elongation complex, but has little effect on stabilizing the initiation complex (*31*).

Strand separation of downstream DNA. Since DNA-dependent RNA polymerases transcribe double-stranded DNA, they must displace the non-template strand from the downstream duplex as it enters the bubble to generate the single-stranded template, thereby functioning as a helicase in addition to a polymerase. Two components of the elongation complex structure, sub-domain H and Helix Y (residues 644-661), appear to be involved in this process. Helix Y is wedged in the fork where the template and non-template strands separate from the downstream duplex, whereas sub-domain H stabilizes the non-template strand of DNA. A bulky amino acid residue, Phe644 at the end of the Y helix, extends outward and stacks on the template base at position *n+1*, the first base pair at the downstream end of the transcription bubble (Fig. 5). Helix Y serves to divert the direction of the non-template strand promoting its separation from the template; this is analogous to the role of the thumb helix, in diverting the direction of the 5′ end of the RNA transcript as it separates from template.

Ever since the first structural studies of the *E. coli* Klenow fragment of Pol 1 (*13, 32*) and continuing through those of substrate complexes with the pol I family of enzymes (*14, 24, 33*), it has remained obscure how DNA polymerase I is able not only to fill single-stranded gaps, but also to displace the RNA primers of Okazaki fragments and synthesize DNA leaving only a nick in the DNA duplex (*34*). Comparison of the structure of T7 RNAP bound to downstream duplex DNA with that of the Pol I family of DNA polymerases provides structural insights into this process. Superposition of the C_α backbones of the palm domains of T7 RNAP and *E. coli* Klenow (KF) fragment aligns the homologous portions of the respective fingers domains, and the downstream duplex DNA bound to T7 RNAP fits well onto the fingers domain of KF (Fig. 5). Helix Y of T7RNAP aligns precisely on helix M of KF which lies between the template and non-template strands. Furthermore, Phe 644 of T7 RNAP that stacks on the last template base of the downstream duplex is identically positioned as Phe 771 of KF. The Phe at this position is conserved in all Pol I family polymerases as a large hydrophobic residue implying that the Pol I family polymerases all use a similar mechanism for binding downstream duplex and separating the two strands. A similar structure is not present in other DNA polymerases, such as the B family of replication polymerases that do not exhibit strand displacement ability.

In the model of Pol I constructed with downstream DNA, the non-template strand departs the Pol I downstream duplex in the direction of the 5′ nuclease domain(*35*), which is responsible for cleaving the Okazaki RNA, while the template strand enters the polymerase active site in the same way as the template strand in Pol I DNA polymerase binary complexes (*24*). In the DNA Pol I binary complexes as in T7RNAP, the template base that will pair with the incoming dNTP is the *n* position and lies in a pocket until the incoming nucleotide arrives to form the ternary complex of enzyme, primer-template and NTP. After nucleotide insertion in DNA Pol I, the template nucleotide *n* now becomes base paired, creating continuous duplex DNA with a nick in the non-template strand between the *n* base pair and *n+1* base pair, which is the first in the downstream duplex.

Inhibition of T7 RNAP transcription by T7 lysozyme. Biochemical data implied that T7 lysozyme may inhibit transcription by preventing T7 RNAP from undergoing the transition from the initiation to the elongation phase (*36*) though previous structural data implied that lysozyme binding may additionally alter the site of catalysis by repositioning the C-terminus (*10*). T7 lysozyme negatively regulates T7 RNAP by binding to it either during or before the initiation phase of transcription, in which case only short abortive transcripts are made, but T7 RNAP is largely unaffected by T7 lysozyme once it has entered the elongation phase except that there is reduced synthesis past pause sites containing an RNA helical hairpin (*6, 8*). The co-crystal structure of T7 RNAP and lysozyme shows the polymerase in largely its initiation phase conformation (except for the extreme C-terminus) with lysozyme bound to the C-terminal domain at some distance from the promoter and catalytic sites (*9, 15*). The T7 lysozyme structure was modeled onto the elongation conformation of T7 RNAP after superposition of the lysozyme and elongation complex palm domain structures. The lysozyme fits well over most of its interaction surface, and the specificity loop moves significantly closer to the lysozyme in its elongation complex conformation. Comparison of the structure of the lysozyme binding sites on T7 RNAP in the initiation and elongation conformations reveals no striking differences. Lysozyme bound to the elongation conformation, however, would be immediately adjacent to the specificity loop and not far from the tunnel exit. Perhaps the 5′ end of the message and/or the specificity loop make a new interaction that prevents elongation of the transcript beyond 15 nucleotides as is observed biochemically (*8*). We must conclude that our understanding of the structural basis for T7RNAP inhibition by lysozyme is at present incomplete.

Comparison of T7 RNAP with the multi-subunit RNA polymerase. Comparisons of the structures of T7 RNAP and its various substrate complexes with those of the multisubunit DNA dependent RNA polymerase that have been recently determined (*37, 38*), show several similarities as well as a few differences. The T7 RNAP elongation complex has a similar angle (~90°) between the axes of the downstream DNA and the heteroduplex as that observed in the yeast pol II elongation complex (*38*) (although the dihedral angles relative to the primer terminus differ), and the length of the heteroduplex is similar (8 base pairs vs 9 base pairs when comparing post-insertion states). Although there are three

unpaired template bases between the last nucleotide of the heteroduplex and the first observed base pair of downstream duplex of Pol II, they are in the B-form DNA conformation and could be base paired in a true transcription bubble in which case the terminal base pair of the heteroduplex and the first base pair of the downstream duplex would be adjacent as in T7 RNAP. The functional reasons for these similarities in the structures of substrates bound to two nonhomologous RNA polymerase are unclear, but they may be related to common mechanisms of translocation, duplex opening, or access to and/or correct selection of incoming NTP. T7 RNA polymerase has a tunnel-like opening, as do the DNA polymerases, most notably the β-family polymerases (*39*) and the multisubunit RNA polymerases (referred to as the "funnel" in yeast Pol II) (*37*), which provides access for the incoming NTP or dNTP. The large angle between the heteroduplex and downstream DNA allows unhindered access of the incoming nucleotide to the primer terminus. Also analogous in the two polymerases, the binding of downstream DNA to yeast Pol II results in the rotation of a domain, called the flap, that sequesters the DNA, a conformational change that may be functionally similar to the closing of loops in the fingers domain of T7 RNAP around the downstream DNA subsequent to its binding. Such sequestering of downstream DNA upon its binding to either polymerase family seems unlikely to be responsible for processive elongation synthesis (*37*, *38*) since such changes presumably also occur upon formation of the initiation complex.

After superimposing the structure of Taq RNA polymerase complex with promoter DNA (*40*) on that of yeast pol II complexed with downstream DNA (*38*), the bend angle between the upstream DNA on the former and the downstream DNA on the latter can be measured and is again similar between the multisubunit polymerase (110°) and T7 RNAP (100°). The size of the upstream and downstream duplex binding sites are about 2 and 1.5 times larger, respectively, in the multisubunit polymerase than the corresponding interaction sites in T7 RNAP. This difference may be related to the larger energy required to open the 12 nucleotide bubble in the former compared with the 6 nucleotide bubble in the latter. Further, like T7 RNAP, both the bacterial and the yeast RNAPs have a presumed RNA exit tunnel that lies near the 5' terminus of the RNA in the heteroduplex which exists also in the apo-enzymes. However, in the bacterial holoenzyme, which also contains the σ subunit that is responsible for promoter specific initiation, the tunnel is blocked by a domain of σ (*37*). It has been suggested that the transition from the initiation phase to the elongation phase in bacteria is triggered by formation of an RNA transcript that is sufficiently long to displace σ from the tunnel thereby facilitating σ's dissociation from the complex and promoter release (*40*, *41*), a hypothesis yet to be verified. Only when a transcript is long enough to displace σ and pass through the tunnel would processive synthesis commence. Presumably, processive RNA synthesis in the elongation phase results from the RNA transcript being surrounded by protein in both polymerase families. A major difference, then, between the multisubunit and T7 RNAPs is the massive conformational change exhibited by the latter to form an exit tunnel that already pre-exists in the former but is blocked in the initiation phase by σ. The unprecedented conformational dexterity exhibited by T7 RNAP may be a consequence of the limited genome space of the T7 phage, which may impose the requirement for this dual functionality of promoter recognition and tunnel formation by the N-terminal domain.

Conclusion. The crystal structure of a T7RNAP elongation complex shows that the N-terminal domain rearranges from its structure in the initiation complex, which destroys the promoter binding site and creates both a channel that accommodates a 7 base pair heteroduplex as well as a tunnel through which the transcript passes after peeling off the heteroduplex. These features account for the enzyme's processivity in the elongation phase as well as the phenomenon of promoter clearance. The fingers domain forms a binding site for 10 base pairs of downstream DNA whose orientation relative to the upstream promoter seen in previous complexes suggests that the enzyme uses the interaction with upstream and downstream duplex to bend and unwind DNA at the transcription start site and thus facilitate promoter opening. The template and non-template strands entering the active site from the downstream DNA are separated by an a-helix that is also present in the homologous DNA polymerase I and may explain how DNA polymerase I is able to displace the 5′ end of the non-template strand. Comparison of the structural differences between the transcribing RNA polymerase complexes and the homologous DNA polymerase I explains how the additional functional properties exhibited by the RNA polymerase are acquired.

References and Notes

1. P. H. von Hippel, D. G. Bear, W. D. Morgan, J. A. McSwiggen, *Annu Rev Biochem* **53**, 389-446. (1984); P. H. von Hippel, *Science* **281**, 660-5. (1998).
2. A. J. Carpousis, J. D. Gralla, *Biochemistry* **19**, 3245-53. (1980).
3. C. T. Martin, D. K. Muller, J. E. Coleman, *Biochemistry* **27**, 3966-74. (1988).
4. R. A. Ikeda, C. C. Richardson, *Proc Natl Acad Sci U S A* **83**, 3614-8. (1986).
5. S. I. Gunderson, K. A. Chapman, R. R. Burgess, *Biochemistry* **26**, 1539-46. (1987).
6. J. Huang, R. Sousa, *J Mol Biol* **303**, 347-58. (2000).
7. Y. Jia, S. S. Patel, *Biochemistry* **36**, 4223-32. (1997); M. L. Ling, S. S. Risman, J. F. Klement, N. McGraw, W. T. McAllister, *Nucleic Acids Res* **17**, 1605-18. (1989).
8. B. A. Moffatt, F. W. Studier, *Cell* **49**, 221-7. (1987); X. Zhang, F. W. Studier, *J Mol Biol* **269**, 10-27. (1997).
9. D. Jeruzalmi, T. A. Steitz, *Embo J* **17**, 4101-13. (1998).
10. G. M. Cheetham, D. Jeruzalmi, T. A. Steitz, *Nature* **399**, 80-3. (1999).
11. G. M. Cheetham, T. A. Steitz, *Science* **286**, 2305-9. (1999).
12. P. Davanloo, A. H. Rosenberg, J. J. Dunn, F. W. Studier, *Proc Natl Acad Sci U S A* **81**, 2035-9. (1984).
13. D. L. Ollis, P. Brick, R. Hamlin, N. G. Xuong, T. A. Steitz, *Nature* **313**, 762-6. (1985).

14. C. A. Brautigam, T. A. Steitz, *Curr Opin Struct Biol* **8**, 54-63. (1998).
15. G. M. Cheetham, T. A. Steitz, *Curr Opin Struct Biol* **10**, 117-23. (2000).
16. R. A. Ikeda, C. C. Richardson, *J Biol Chem* **262**, 3800-8. (1987).
17. B. He, M. Rong, R. K. Durbin, W. T. McAllister, *J Mol Biol* **265**, 275-88. (1997).
18. E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, *Cell* **89**, 33-41. (1997); N. Korzheva, A. Mustaev, E. Nudler, V. Nikiforov, A. Goldfarb, *Cold Spring Harb Symp Quant Biol* **63**, 337-45. (1998).
19. D. Temiakov *et al.*, *Proc Natl Acad Sci U S A* **97**, 14109-14. (2000).
20. Template-dependent transcription by T7 RNAP was assayed by incorporation of P32 labeled nucleotides into the RNA transcript using the designed DNA substrate. Binding affinity of the designed bubble construct (Fig. 1A) to T7RNAP was determined by gel shift assay and florescence polarization measurement. The apparent Kd was estimated to be 5 μM under the crystallization conditions
21. S. S. Daube, P. H. von Hippel, *Science* **258**, 1320-4. (1992); S. S. Daube, P. H. von Hippel, *Biochemistry* **33**, 340-7. (1994).
22. The transcription bubble substrate was assembled by mixing three oligonucleotides at a molar ratio of template/non-template/RNA of 1.0:1.0:1.1. The annealing solution contained 50 mM, pH 7.0 Tris-HCl, 10 mM Mg acetate, 20 mM NaCl2 and was heated to 75° for 5 min. The mixture was then slowly cooled to 20°C over a period of 2 hours. The transcription elongation complex was then assembled by mixing the annealed substrate with equimolar T7RNAP.
23. Crystals of the T7 RNAP transcription elongation complex were obtained by vapor diffusion; 250 μM of elongation complex solution was mixed with an equal volume of a well solution containing 50 mM Tris-HCl , pH 7.0, 200 mM Li2SO4, 6% PEG 8000, 10 mM Mg acetate and 5 mM dithiothreitol. The crystals were flash-frozen in liquid propane prior to data collection. Selenomethionine substituted T7RNAP elongation complex was also prepared for MAD experiments (*44*). Native data were collected at Brookhaven National Laboratory, beamline X25, and single wavelength anomalous dispersive (SAD) data were collected at APS, beamline ID19. Data were reduced with DENZO and Scalepack (*45*). The scaled data have Rlinear 6.5 and 7.2% for the native and SAD data set, respectively. The derivative data were obtained with a crystal in monoclinic form. The data was moved to the orthorhombic form for refinement. The crystals are in space group C2221 with unit cell dimension a=144, b= 145,c=144 Å, 90°, 90°, 90°. The structure was determined by combining phases from Se SAD and molecular replacement using the C-terminal domain (residues 300-883) of the T7RNAP initiation complex with the programs Amore (*46*) and CNS (*47*). Density modification was applied to reduce phase error. The coordinates were refined against the 2.1 Å resolution native data to an Rfree of 28%.
24. J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **391**, 304-7. (1998).
25. N. M. Stano, S. S. Patel, *J Mol Biol* **315**, 1009-25. (2002); L. G. Brieba, R. Sousa, *Biochemistry* **40**, 3882-90. (2001).
26. P. A. Bullough, F. M. Hughson, J. J. Skehel, D. C. Wiley, *Nature* **371**, 37-43. (1994).
27. J. Kuriyan, M. O'Donnell, *J Mol Biol* **234**, 915-25. (1993).
28. V. A. Bloomfield, D. M. Crothers, I. Tinoco, Jr., *Nucleic acids-structures, properties, and functions* (Universisty Science Books, Sausalito, CA, 2000).
29. F. M. Richards, *Annu Rev Biophys Bioeng* **6**, 151-76. (1977).
30. G. A. Diaz, M. Rong, W. T. McAllister, R. K. Durbin, *Biochemistry* **35**, 10837-43. (1996); Y. Jia, A. Kumar, S. S. Patel, *J Biol Chem* **271**, 30451-8. (1996); A. Ujvari, C. T. Martin, *J Mol Biol* **273**, 775-81. (1997).
31. V. Gopal, L. G. Brieba, R. Guajardo, W. T. McAllister, R. Sousa, *J Mol Biol* **290**, 411-31. (1999).
32. P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, T. A. Steitz, *Proc Natl Acad Sci U S A* **85**, 8924-8. (1988).
33. S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **391**, 251-8. (1998); S. Doublie, T. Ellenberger, *Curr Opin Struct Biol* **8**, 704-12. (1998).
34. A. Kornberg, T. Baker, *DNA replication* (W H Freeman & Co.; 2nd edition, 1991).
35. U. K. Urs, R. Murali, H. M. Krishna Murthy, *Acta Crystallogr D Biol Crystallogr* **55**, 1971-7. (1999).
36. J. Huang, J. Villemain, R. Padilla, R. Sousa, *J Mol Biol* **293**, 457-75. (1999); A. Kumar, S. S. Patel, *Biochemistry* **36**, 13954-62. (1997).
37. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1876-82. (2001).
38. P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1863-76. (2001).
39. M. C. Franklin, J. Wang, T. A. Steitz, *Cell* **105**, 657-67. (2001).
40. K. S. Murakami, S. Masuda, E. A. Campbell, O. Muzzin, S. A. Darst, *Science* **296**, 1285- 90. (2002).
41. D. G. Vassylyev *et al.*, *Nature* **417**, 712-9. (2002); S. S. Daube, P. H. von Hippel, *Proc Natl Acad Sci U S A* **96**, 8390-5. (1999).
42. J. A. Christopher. (Jon A Christopher, Texas A&M University, College Station, TX 77843- 2128, 1997).
43. M. Carson, in *Methods in Enzymology* R. M. S. a. C. W. Carter, Ed. (Academic Press, 1997), vol. 277, pp. 493-505.
44. Doublie, S, in *Methods in Enzymology* R.M.S.a.C.W. Carter, Ed.) (Academic Press, 1997), vol. 277, pp. 523-530.
45. Z. Otwinowski and W. Minor, in *Methods in Enzymology* R.M.S. a. C.W. Carter, Ed. (Academic Press, 1997), vol. 277, pp. 307-326.
46. M. Navarro, G.A. Cross, E. Wirtz. *EMBO J* **18**, 2265-72. (1999).
47. A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S., Jiang, J. M. Kuszewski, M. Hilges, N. S. Pannu, R. J., Read, L. M.

Rice, T. Simonson, G. L. Warren. *Acta Crystallogr* **D54**, pp. 905-21. (1998).

48. Data were collected at two synchrotron sources for this work, ID-19 at Argonne National Labs (APS) and X25 at National Synchrotron Light Source (NSLS). Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source, was supported by the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract No. W-31-109-ENG-38. Research carried out in part at the National Synchrotron Light Sources, Brookhaven National Laboratory, is supported by the U.S. Department of Energy, Division of Materials Sciences and Division of Chemical Sciences, under Contract No. DE-AC02-98CH10886. We wish to thank Michael Becker (X25) and Andrzej Joachimiak (ID-19) for their support during beamline data collections. The authors with to thank Wm. Kennedy, C. Joyce, and S. Kamtekar for many helpful discussions and critical reading of the manuscript. We thank D. Crothers for help with the thermodynamic calculations. The coordinates for the T7RNAP elongation complex are being deposited in the PDB. This work was supported by NIH grant GM57510 to TAS.

16 August 2002; accepted 11 September 2002
Published online 19 September 2002;
10.1126/science.1077464
Include this information when citing this paper.

Fig. 1. Substrates in the T7RNAP elongation complex. (A) The substrate construct co-crystallized with T7 RNAP consisted of 30 nucleotides each of template DNA (blue) and non-template DNA (green) that are complementary (except for a central 11 nucleotides) and a 17 nucleotide RNA (red) whose 3′ 10 nucleotides are complementary to the template DNA. The nucleotide that templates the nascent NTP is numbered 0 and nucleotides upstream are given negative numbers and downstream given positive numbers. The portions of RNA and DNA that are not visible in the map are outlined by dashed lines. (B) A portion of the composite omit electron density map corresponding to the substrate is contoured at 1.1 σ. Figure 1 and Figure 4 were made with program SPOCK (*42*).

Fig. 2. Comparison of the structures of the T7 RNAP initiation and elongation complexes. The initiation complex (A) and elongation complex (B) have been orientated equivalently by superimposing their palm domains. Helices are represented by cylinders and ß-strands by arrows. The corresponding residues in the N-terminal domains of the two complexes that undergo major refolding are colored in yellow, green and purple, and the C-terminal domain (residues 300-883) is colored in gray. The template DNA (blue), non-template DNA (green) and RNA (red) are represented with ribbon backbones. The proteolysis susceptible region (residues 170-180) is a part of sub-domain H (green) in the elongation complex and has moved more than 70 Å from its location in the initiation complex. The specificity loop (brown) recognizes the promoter during initiation, and contacts the 5′ end of RNA during elongation while the intercalating hairpin (purple) opens the upstream end of the bubble in the initiation phase and is not involved in elongation. The large conformational change in the N-terminal region of T7RNAP facilitates promoter clearance. (C) The DNA and trinucleotide of RNA seen in the structure of the initiation complex docked to the downstream DNA of the elongation complex shows a 100° bend between upstream and downstream segments. (D) The DNA and 7 nucleotides of RNA observed in the structure of the elongation complex shows a decreased angle of bending between the base paired upstream and downstream segments. Figures (E), (F) and (G) show the three different conformational changes – helix formation, rigid body movement and refolding – undergone in the transition between initiation and elongation. This figure and Figures 3, 5, 6 were made with the program *Ribbons* (*43*).

Fig. 3. Views of the transcription bubble. (A) Global view of the elongation complex with a box outlining the active site region which is enlarged in (B), (C), and (D) with the thumb in yellow green, subdomain H in green, specificity loop in yellow and helix Y in red. (B) Conformational changes of the thumb and the specificity loop. The thumb domain as observed in the initiation complex (grey) has rotated about 15° in the elongation complex (green) and assists in the separation of the RNA transcript from the template DNA. The position of the specificity loop in the initiation complex (yellow) blocks the exit of RNA and has moved in the elongation complex (brown) to open the exit tunnel and interact with the exiting RNA. The three basepairs of heteroduplex in the initiation complex (grey) superimposes on that of the elongation complex. (C) Interactions of the transcription bubble and heteroduplex in the elongation complex with domain H (green and red) and specificity loop (brown). Proteolytic cuts within the red loop in subdomain H reduce elongation synthesis (*3*, *16*). Thumb σ-helix (yellow) and σ- helix Y (orange) are analogously involved in strand separation. (D) Shown are side-chains from subdomain H (green), the specificity loop (brown) and the thumb that interact with the single-stranded 5′ end of the RNA transcript and facilitate its separation from the template.

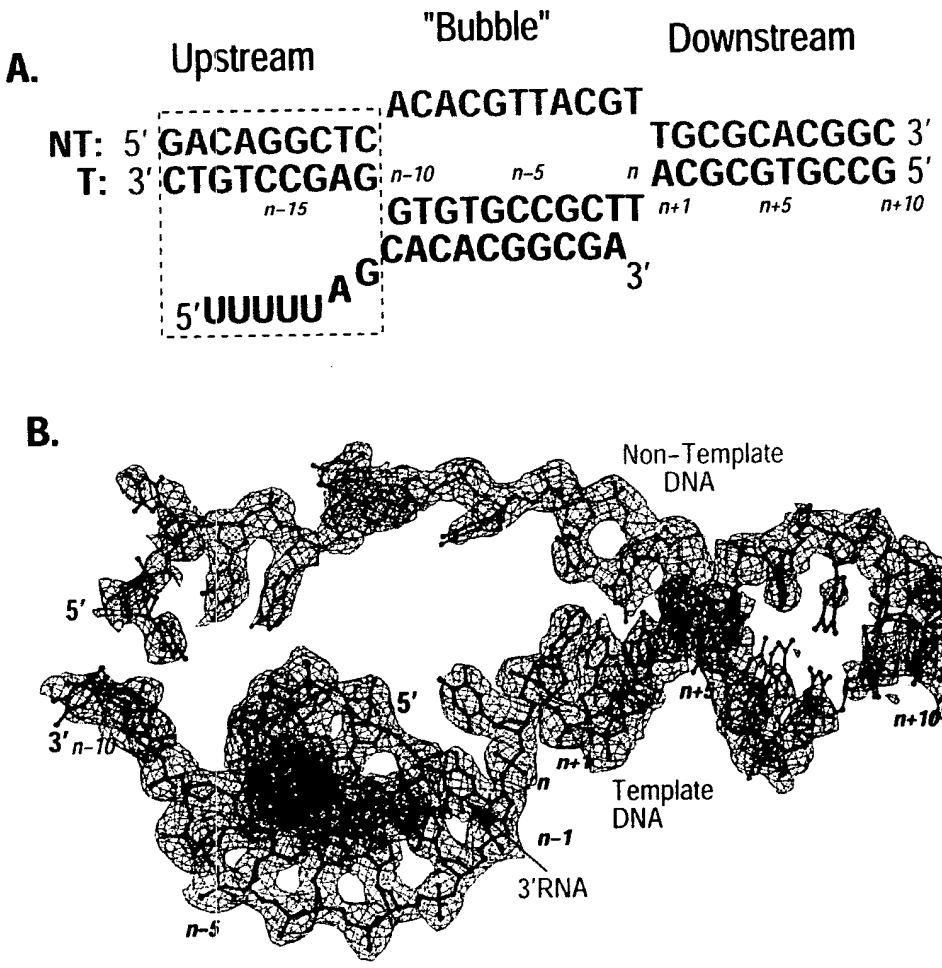
Fig. 4. Formation of the RNA exit tunnel and upstream DNA binding site in the elongation complex. The solvent contact surface for the initiation complex conformation of T7 RNAP (A) with the observed upstream promoter DNA and heteroduplex along with the downstream DNA modeled from the elongation complex. The thumb domain has been removed from both A and B to allow a view into the heteroduplex binding site. The elongation complex (B) shows the disappearance of the promoter DNA binding site, the formation of a new channel that binds to heteroduplex and upstream DNA nonspecifically and a tunnel through which the transcript (red) is exiting. (C) and (D) are rotated by 180° about a vertical axis and the appearance of tunnel that contains the 5′ end of an RNA transcript (red) in the elongation complex (D). The positive electrostatic potential is blue and the negative is red. The two complexes have been oriented identically by superposition of their palm domains.

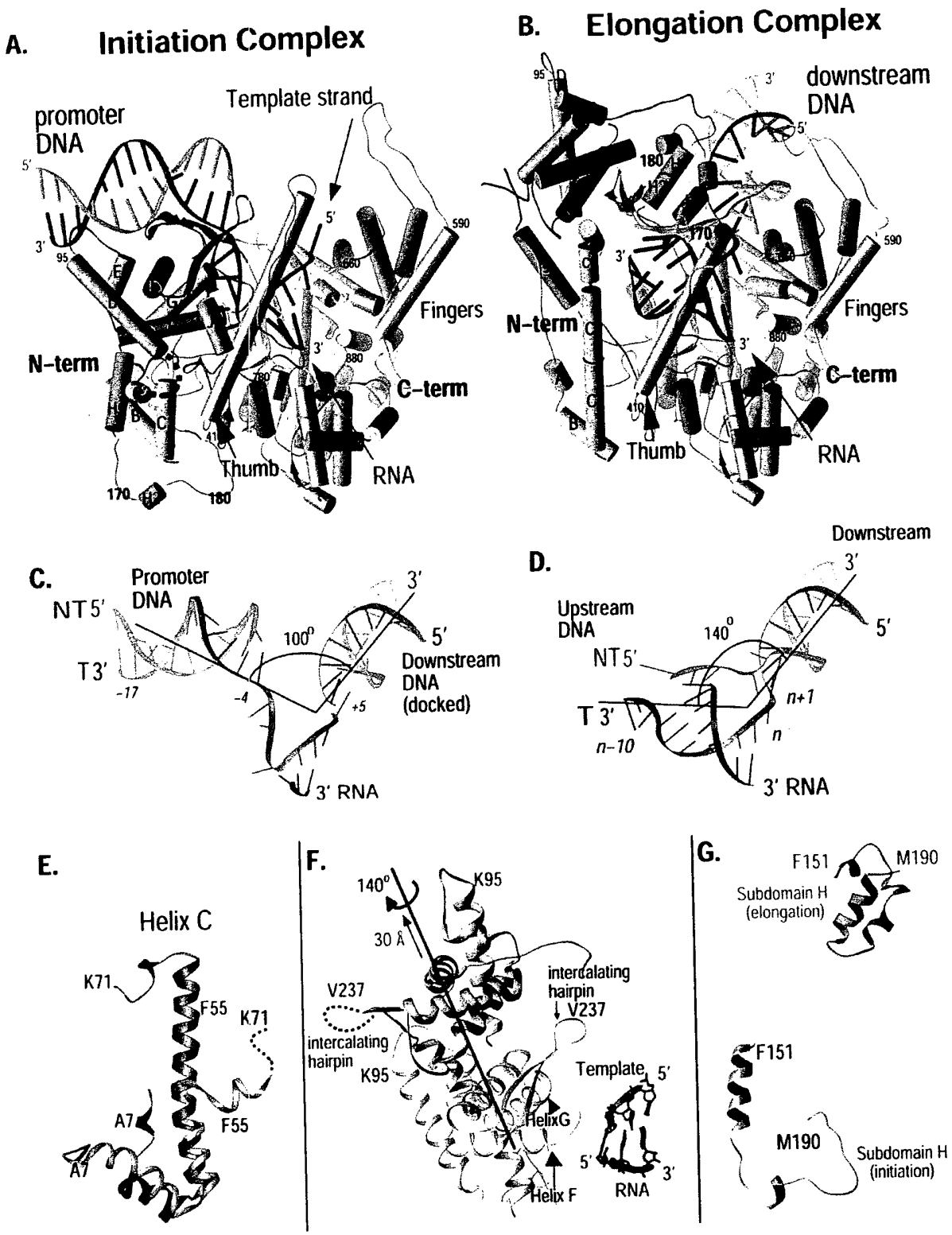
Fig. 5. The T7 RNAP fingers domain (grey) bound to the downstream DNA superimposed on the corresponding part of the Klenow fragment fingers domain (yellow). The template DNA (blue) is redirected and separated from the non-template DNA (green) by an α-helix Y and F644 in T7 RNAP. A corresponding α-helix M and Phe are found in the in Klenow fragment as are other portions of the downstream DNA binding site. Part of the RNA transcript is shown in red.

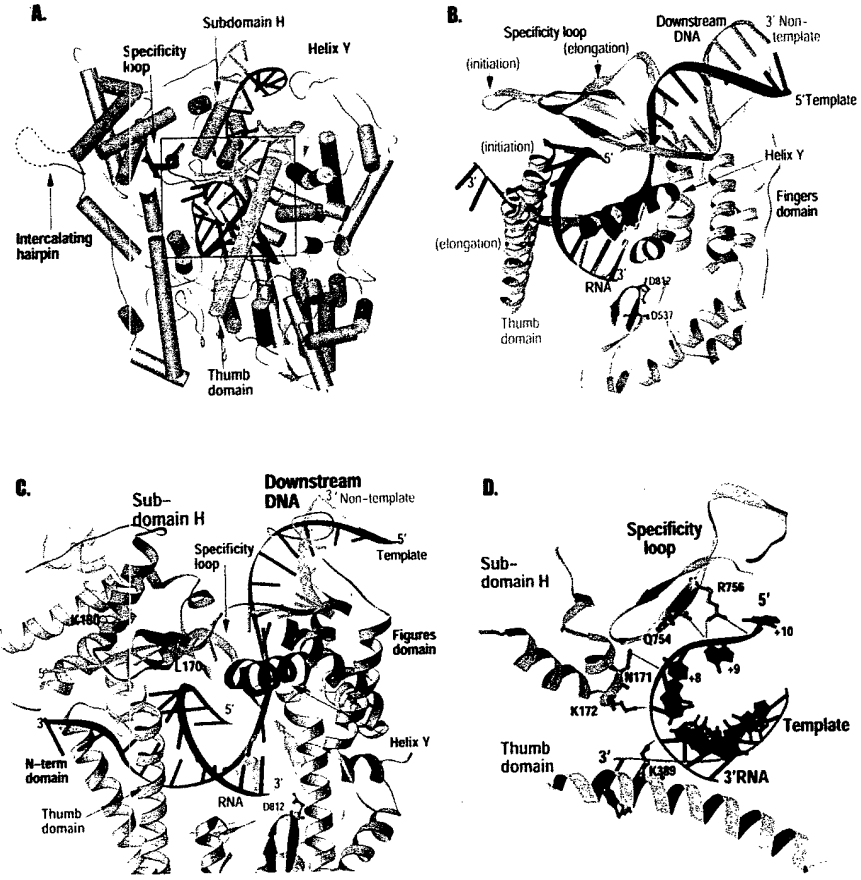
Table 1. Summary of crystallographic analysis

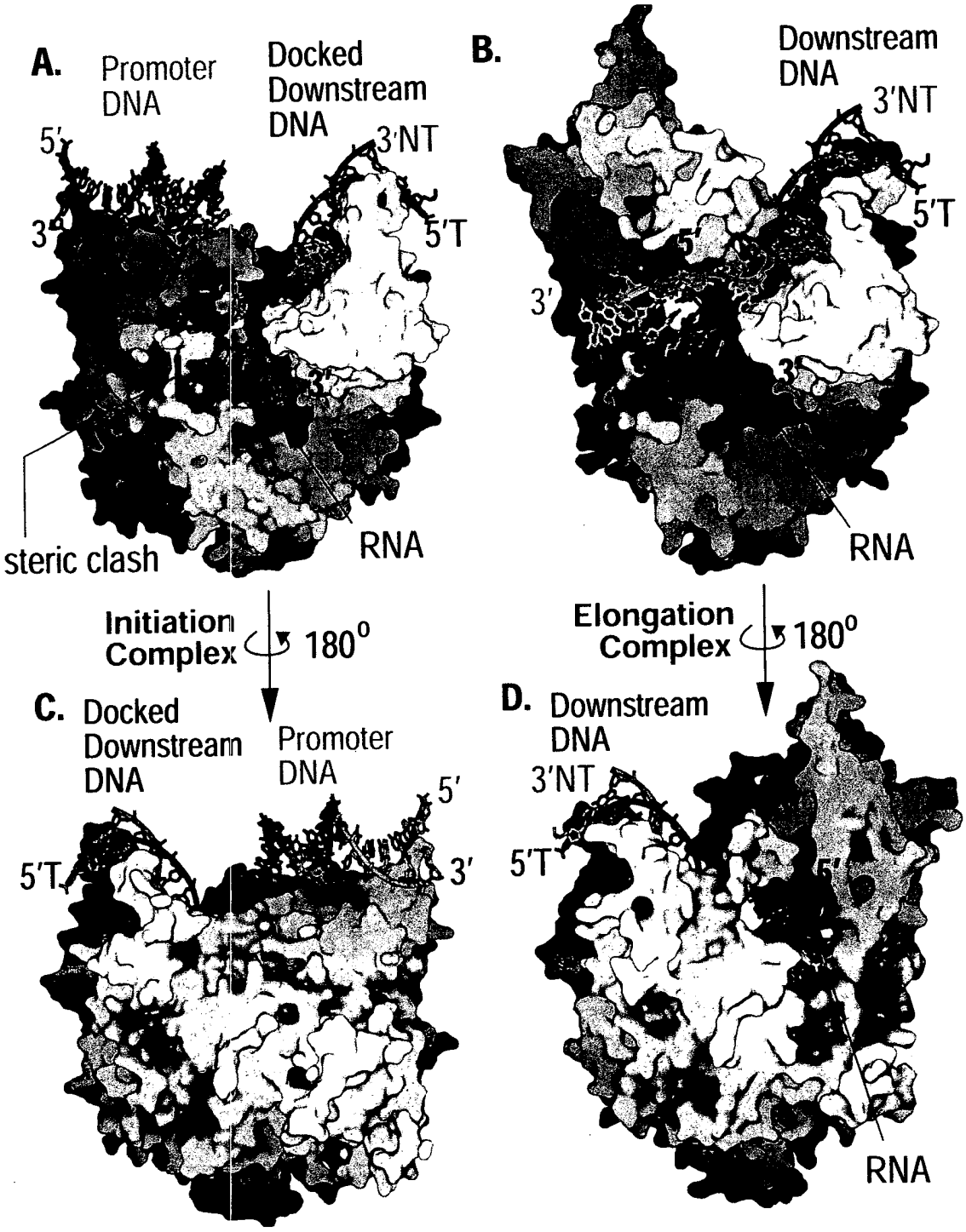
Data collection and SAD analysis	Native1	Native2	SeMet(SAD)
Wavelength (Å)	0.98	0.98	0.979
Resolution (Å)	2.1	2.2	2.9
Completeness (overall/last shell, %)	85.0 (35.6)	89.2 (55.4)	100(100)
Space group	$P2_1$ (2 copies/a.u)	$C222_1$ (1 copy/a.u.)	$P2_1$ (2copies/a.u.)
Unit cell (Å)	100.7 144.8 101.2 90.0 90.6 90.0	142.9 145.5 145.6 90.0 90.0 90.0	100.8 144.2 102.0 90.0 91.1 90.0
No. of sites			50
R_{linear}	6.5	8.2	7.2
Phasing power (acentric reflections)			1.2
Solvent-flipping density modification SAD FOM			0.62
Structure refinement			
Resolution (Å)	40-2.09 *(Fo > 2s|F|)*		
Rfactor/Rfree	24.1/27.3		
No. protein residues/nucleic acid/ water	862 amino acids (missing residues 1, 233-240, 363-374) 47 nucleotides, 190 water molecules.		

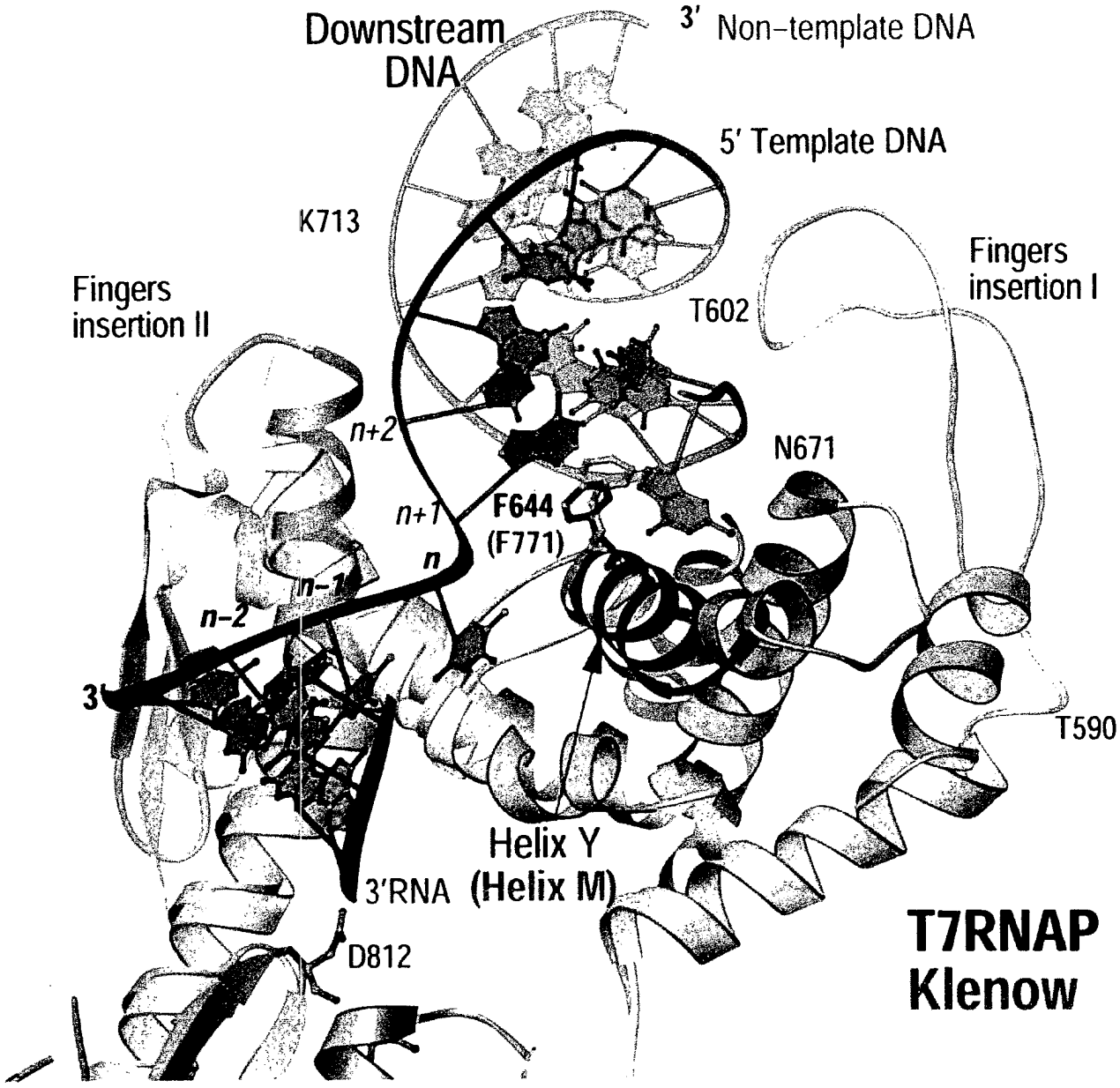
R_{linear}=| *I*-<*I*>|/*I*, where *I* is the observed intensity and <*I*> is the average intensity for multiple measurements of reflections.
Phasing power= r.m.s.(|*Fph*|/*E*), where *E* is the residual lack of closure.
FOM, figure of merit.
Rfree, R factor calculated using the test data set that was excluded from the refinement.
R=(|(*Fp*|-|*Fc*|)/|*Fp*|, where *Fp* and *Fc* are the observed and calculated structure factors.


A.
Upstream
"Bubble"
Downstream
ACACGTTACGT
NT: 5' GACAGGCTC
T: 3' CTGTCCGAG
n-15
n-10
n-5
n
n+1
n+5
n+10
TGCGCACGGC 3'
ACGCGTGCCG 5'
GTGTGCCGCTT
CACACGGCGA
3'
5'UUUUU
A
G
B.
Non-Template DNA
5'
3'
3' n-10
5'
n+5
n+10
5'
n+1
n
Template DNA
n-1
3'RNA
n-5


A.
Initiation Complex
promoter DNA
Template strand
5'
3'
95
590
660
880
780
410
170
180
Fingers
N-term
C-term
Thumb
RNA
B.
Elongation Complex
downstream DNA
95
3'
5'
180
170
590
880
410
Fingers
N-term
C-term
Thumb
RNA
C.
Promoter DNA
NT 5'
T 3'
3'
5'
100°
-17
-4
+5
Downstream DNA (docked)
3' RNA
D.
Downstream
3'
5'
Upstream DNA
NT 5'
T 3'
140°
n+1
n
n-10
3' RNA
E.
Helix C
K71
F55
K71
A7
F55
A7
F.
140°
K95
30 Å
intercalating hairpin
V237
V237
intercalating hairpin
K95
Template
5'
HelixG
5'
3'
RNA
Helix F
G.
F151
M190
Subdomain H (elongation)
F151
M190
Subdomain H (initiation)


A.
Subdomain H
Specificity loop
Helix Y
Intercalating hairpin
Thumb domain
B.
Specificity loop
(elongation)
Downstream DNA
3' Non-template
(initiation)
5'Template
(initiation)
Helix Y
Fingers domain
(elongation)
RNA
Thumb domain
C.
Sub-domain H
Downstream DNA
3' Non-template
Specificity loop
5' Template
K180
L170
Figures domain
Helix Y
N-term domain
RNA
Thumb domain
D.
Specificity loop
Sub-domain H
R756
5'
Q754
+10
N171
+8
+9
K172
Template
Thumb domain
3'RNA


A.
Promoter DNA
Docked Downstream DNA
5′
3′
3′NT
5′T
steric clash
RNA
Initiation Complex
180°
B.
Downstream DNA
3′NT
5′T
5′
3′
3′
RNA
Elongation Complex
180°
C.
Docked Downstream DNA
Promoter DNA
5′
5′T
3′
D.
Downstream DNA
3′NT
5′T
5′
RNA


Downstream DNA
3' Non-template DNA
5' Template DNA
K713
Fingers insertion II
T602
Fingers insertion I
n+2
N671
n+1
F644 (F771)
n
n-1
n-2
3'
T590
Helix Y (Helix M)
3'RNA
D812
T7RNAP Klenow

Reference molecule: T7 1 - 883 (883 aa) Homology

Sequence 2:	C21	1 - 883	(883 aa)	98%
Sequence 3:	T3	1 - 884	(884 aa)	88%
Sequence 4:	K11	1 - 906	(906 aa)	82%
Sequence 5:	SP6	1 - 874	(874 aa)	48%

Parameters set: **Mismatch** = 3; Open Gap = 5; **Extend Gap** = 2

Conservative Amino Acid Changes Permitted

```
                10        20        30        40        50        60        70        80        90
                *         *         *         *         *         *         *         *         *
T7      MNTI-NIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRKMFERQLKAGEVADNAAAKPLITTLLPKMIARINDWFEEVKAKRGKR
C21     MNTI-NIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRKMFERQLKAGEVADNAAAKPLITTLLPKMIARINDWFEEVKAKRGKR
T3      MNiIeNIeKNDFSeIELAAIPFNTLADHYGsaLAkEQLALEHESYElGErRFlKMlERQaKAGEiADNAAAKPLlaTLLPKlttRIveWlEEyasKkGrk
K11     MNal-NIgrNDFSeIELAAIPyNiLseHYGdqaAREQLALEHEaYElGrqRFlKMlERQvKAGEfADNAAAKPLvlTLhPqltkRIdDWkEEqanaRGKk
SP6     Mq------------dLhAI---------------QLqLEeEmfngGirRFeadqqRQiaAGsesDtAwnrrLlseLiapMaegIqaykEEyegKkGra

                                  110       120       130       140       150       160       170
                                  *         *         *         *         *         *         *
T7      PTAFQF----------------LQEI---KPEAVAYITIKTTLACLTSADNTTVQAVASAIGRAIEDEARFGRIRDLEAKHFKKNVEEQLNKRVGHVYK
C21     PTAFQF----------------LQEI---KPEAVAYITIKTTLACLTSvDNTTVQAVASAIGRAIEDEAaFGRIRDLEAKHFKKNVEEQLNKRVGHVYK
T3      PsAyap----------------LQll---KPEAsAfITlKviLAsLTStnmTTiQAaAgmlGkAIEDEARFGRIRDLEAKHFKKhVEEQLNKRhGqVYK
K11     PrAyypikhqvaselavsmgaevLkEkrgvssEAiAllTIKvvLgnahrplkghnpAVsSqlGkAlEDEARFGRIReqEAayFKKNVadQLdKRVGHVYK
SP6     PrAlaF----------------LQcv---enEvaAYITmKvvmdmLnt--daTlQAiAmsvaerIEDqvRFskleghaAKyFeK-VkksLkasrtksYr

                  190       200       210       220                 230       240       250       260
                  *         *         *         *                   *         *         *         *
T7      KAFMQVVEADM-LSKGLLGGEAWSSWHKEDSIHVGVRCIEML--------------IESTGMVSLHRQNAGVVGQDSETI---ELAPEYAEAIATRAGA
C21     KAFMQVVEADM-LSKGLLGGEAWSSWHKEDSIHVGVRCIEML--------------IESTGMVSLHRQNAGVVGQDSETI---ELAPEYAEAIATRAGA
T3      KAFMQVVEADM-igrGLLGGEAWSSWdKEttmHVGiRlIEML--------------IESTGlVeLqRhNAGnaGsDhEal---qLAqEYvdvlAkRAGA
K11     KAFMQVVEADM-iSKGmLGGdnWaSWktdegmHVGtkllElL--------------IEgTGlVemtknkmadgsdDvtsmqmvqLAPafvEllskRAGA
SP6     hAhnvaVvAeksvaekdadfdrWeaWpKEtqlqiGttllEiLegsvfyngepvfmramrtyGgktiyylqt------SEsv---------ggwIsafkeh

                270       280       290       300       310       320       330       340                 350
                *         *         *         *         *         *         *         *                   *
T7      LAGISPMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPV-------EDIP
C21     LAGISPMFQPCVVPPKPWTGITGGGYWANGRRPLALVsTHSKKALMRYEDVYMPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPV-------EDIP
T3      LAGISPMFQPCVVPPKPWvaITGGGYWANGRRPLALVRTHSKKgLMRYEDVYMPEVYKAvNlAQNTAWKINKKVLAVvNeIvnWKnCPV-------aDIP
K11     LAGISPMhQPCVVPPKPWvetvGGGYWsvGRRPLALVRTHSKKALrRYaDVhMPEVYKAvNlAQNTpWKvNKKVLAVvNeIvnWKHCPV-------gDvP
SP6     vAqlSPayaPCViPPrPWrtpfnGGfhtekvasrirlvkgnrehvrkltqkqMPkVYKAINalQNTqWqINKdVLAVieevir-----ldlgygvpsfkP

                  360       370                 380       390       400       410       420       430
                  *         *                   *         *         *         *         *         *
T7      AIEREELPMKPEDIDMNP-------EA--------LTAWKRAAAAVYRKDKARKSRRISLEFMLEQANKFANHKAIWFPYNMDWRGRVY--AVSMFNPQG
C21     AIEREELPMKPEDIDtNP-------EA--------LTAWKRAAAAVYRKDKARKSRRISLEFMLEQANKFANHKAIWFPYNMDWRGRVY--AVSMFNPQG
T3      slERqELPpKPdDIDtNe-------aA--------LkeWKkAAAgiYRlDKARvSRRISLEFMLEQANKFAskKAIWFPYNMDWRGRVY--AVpMFNPQG
K11     AIEREELPprPdDIDtNe-------vA--------rkAWrkeAAAVYRKDKARqSRRcrcEFMvaQANKFANHKAIWFPYNMDWRGRVY--AVSMFNPQG
SP6     lIdkEnkPanPvpvefqhlrgrelkEmlspeqwqqfinWKgecArlYtaetkRqSksaavvrMvgQArKysafesIyFvYaMDsRsRVYvqsstl-sPQs
```

```
                440       450       460       470       480       490       500       510       520       530
                 *         *         *         *         *         *         *         *         *         *
T7      NDMTKGLLTLAKGKPI-GKEGYYWLKIHGANCAGVDKVPFPERIKFI--EENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYA---GVQHHGLSYNC
C21     cDMTKGpwTLAKGKPI-GKEGYYWLKIHGANCAGVDKVPFPERIKFI--EdNHdNIMrCAKSPLENTWWAEQDSPFCFLAFCFEYA---GVQHHGLSYNC
T3      NDMTKGLLTLAKGKPI-GeEGfYWLKIHGANCAGVDKVPFPERIaFI--EkhvddIlACAKdPinNTWWAEQDSPFCFLAFCFEYA---GVtHHGLSYNC
K11     NDMTKGsLTLAKGKPI-GldGfYWLKIHGANCAGVDKVPFPERIKFI--EENegNIlAsAadPLnNTWWtqQDSPFCFLAFCFEYA---GVkHHGLnYNC
SP6     NDlgKaLLrfteGrPvnGvEalkWfcInGANlwGwDKktFdvRvsnvldEEfqdmcrdiAadPLtfTqWAkaDaPyeFLAwCFEYAqyldlvdeGradef

                        540       550       560       570       580       590       600       610       620
                         *         *         *         *         *         *         *         *         *
T7      --SLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNEILQADAINGTDNEVVTVTDENTGEISEKVKLG--TKALAGQWLAYGVT
C21     --SLPLAFDGSCSGIQHFSAMLlDEVGGRAVNLLPSETVQDIYGIVAKKVNvILQeDvINGTDNEVVTVTDENTGEISEKVKLG--TKALAGQWLAYGVT
T3      --SLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAqKVNEILkqDAINGTpNEmiTVTDkdTGEISEKlKLG--TstLAqQWLAYGVT
K11     --SLPLAFDGSCSGIQHFSAMLRDsiGGRAVNLLPSdTVQDIYkIVAdKVNEvLhqhAvNGsqtvVeqiaDkeTGEfhEKVtLG--esvLAaQWLqYGVT
SP6     rthLPvhqDGSCSGIQHySAMLRDEVGakAVNLkPSdapQDIYGaVA---qvvikknAlymdaddatTfT---sGsvt---lsGtelrAmAsaWdsiGiT

                630       640       650                 660                 670       680       690       700
                 *         *         *                   *                   *         *         *         *
T7      RSVTKRSVMTLAYGSKEFGFRQQVLEDT-------IQPAIDSGKG------------LMFTQPNQAAGYMAKLIWESVSVTVVAAVEAMNWLKSAAKLLA
C21     RSVTKRSVMTLAYGSKEFGFRQQVLEDT-------IQPAIDSGKG------------LMFTQPNQAAGYMAKLIWEaVSVTVVAAVEAMNWLKSAAKLLA
T3      RSVTKRSVMTLAYGSKEFGFRQQVLdDT-------IQPAIDSGKG------------LMFTQPNQAAGYMAKLIWdaVSVTVVAAVEAMNWLKSAAKLLA
K11     RkVTKRSVMTLAYGSKEslvRQQVLEDT-------IQPAIDnGeG------------LMFThPNQAAGYMAKLIWdaVtVTVVAAVEAMNWLKSAAKLLA
SP6     RSlTKkpVMTLpYGStrltcResVidyivdleekeaQkAvaeGrtankvhpfeddrqdylT-PgaAynYMtaLIWpSiSevVkApivAMkmirqlArfaA

                710       720       730       740       750       760       770       780       790       800
                 *         *         *         *         *         *         *         *         *         *
T7      AEVKDKKTGEILRKRCAVHWVTPDGFPVWQEYKKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVWAHEKYGIESF
C21     AEVKDKKTGEILRKRCAVHWVTPDGFPVWQEYKKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVWAHEKYGIESF
T3      AEVKDKKTkEILRhRCAVHWtTPDGFPVWQEYrKPlQkRLdmiFLGQFRLQPTINTlKDSgIDAHKQESGIAPNFVHSQDGSHLRmTVVyAHEKYGIESF
K11     AEVKDKKTkEvLRKRCAiHWVTPDGFPVWQEYrKqnQaRLkLvFLGQanvkmTyNTgKDSEIDAHKQESGIAPNFVHSQDGSHLRmTVVhAnEvYGIdSF
SP6     ----------krneglmytlPtGFileQkimatemlRvrtclmGdikm--slqvetDivdeAammga-aAPNFVHghDaSHLilTVcelvdK-GvtSi

                810       820       830       840       850       860       870       880
                 *         *         *         *         *         *         *         *
T7      ALIHDSFGTIPADAANLFKAVRETMVDTY-ESCDVLADFYDQFADQLHESQLDKMPALPAKGNLNLRDILESDFAFA
C21     ALIHDSFGTIPrDAANLFKAVRETMVDTY-ESCDVLADFYDQFADQLHESQLDKMPALPAKGNLNLqDILkSDFAFA
T3      ALIHDSFGTIPADAgkLFKAVRETMViTY-EnnDVLADFYsQFADQLHEtQLDKMPpLPkKGNLNLqDILkSDFAFA
K11     ALIHDSsGTIPADAgNLFKAVRETMVkTY-EdnDViADFYDQFADQLHESQLDKMPAvPAKGdLNLRDILESDFAFA
SP6     AvIHDSFGThadntltLrvAlkggMVamYidgnalgklleehevrwmvdtgie----vPeqGefdLneImdSeyvFA
```